6/

REPORT RESUMES

ED 003 044 24
REPRESENTATIVE ORDERING AND SELECTION OF VARIABLES, VOLUMES A AND B--STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA, PART 3.
BY- BARGMANN, ROLF E.
VIRGINIA POLYTECHNIC INST., BLACKSBURG
REPORT NUMBER CRP-1132-PT. A-B PUB DATE JUN 62
EDRS PRICE MF-$0.54 HC-$13.60 340P.

DESCRIPTORS- *EDUCATIONAL RESEARCH, DATA ANALYSIS, STATISTICAL STUDIES, COMPUTER PROGRAMS, TEST CONSTRUCTION, *MATHEMATICAL MODELS, *DISCRIMINANT ANALYSIS, *ANALYSIS OF VARIANCE, VIRGINIA, BLACKSBURG

THE STUDIES EMBODIED IN THIS REPORT PROPOSE SOME STATISTICAL METHODS OF ORDERING AND ATTAINING RELEVANCY TO HELP THE EDUCATIONAL RESEARCHER CHOOSE AMONG SUCH VARIABLES AS TESTS AND BEHAVIOR RATINGS. CONSTRUCTION OF A MODEL FOR THE ANALYSIS OF CONTINGENCY TABLES, DETERMINATION OF THE MOST APPROPRIATE ORDERING PRINCIPLE IN STEP-DOWN ANALYSIS FOR THE QUALITATIVE INTERPRETATION OF STATISTICAL RESULTS, AND DEVELOPMENT OF STATISTICAL TESTS TO INTERPRET THE RESULTS WERE ATTEMPTED. PARTIAL CORRELATION AND PARTIAL ASSOCIATION CONCEPTS ARE ANALYZED TO DETERMINE PARAMETERS FOR GENERALIZING CLASSIFICATION OF DEPENDENCE PATTERNS. PROCEDURES WERE DEVELOPED FOR FACTOR IDENTIFICATION AND COMPARED WITH LATENT CLASS ANALYSIS PROCEDURES CURRENTLY IN USE. SUBJECT-TEST INTERACTION MODELS DESIGNED THROUGH OTHER RESEARCH WERE USED FOR COMPARISON. VARIANCE-COVARIANCE OR CORRELATION MATRIX PATTERNS RESULTING FROM EACH ASSUMPTION CONCERNING THE NATURE OF SUBJECT-TEST INTERACTIONS WERE PRESENTED. STATISTICAL TESTS WERE CONSTRUCTED FOR DETERMINING THE PLAUSIBILITY OF EACH PATTERN AND FOR PATTERN DIFFERENCES. SUBJECT-TEST INTERACTION MODELS WERE IDENTIFIED THROUGH THE DEVELOPMENT OF STATISTICAL METHODS, AND THESE METHODS WERE IN TURN TRANSLATED INTO OPERATIONAL PROCEDURES. IT WAS CONCLUDED THAT STATISTICAL INTERPRETATION OF SIGNIFICANT GROUP DIFFERENCES IS POSSIBLE USING DISCRIMINANT FUNCTIONS AND SUBSEQUENT STEP-DOWN ANALYSIS. ANY DIFFERENCES DUE TO LATENT ARTIFICIAL VARIABLES INSTEAD OF THE OBSERVED VARIABLES MAY BE DETERMINED THROUGH FACTOR AND DISCRIMINATION ANALYSES. OTHER ENTRIES IN THIS SERIES ARE ED 003 045 AND ED 003 046. (HS)

Check Green
Louis M. Heil

not yet Approved

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

REPRESENTATIVE ORDERING AND SELECTION OF VARIABLES

by

Rolf E. Bargmann

Part A

Final Report (Part III, according to Proposal, Appendix A) of Cooperative Research Project No. 1132, "Statistical Models for the Evaluation and Interpretation of Educational Criteria", U. S. *Office of Education, Department of Health, Education,* and Welfare.

Virginia Polytechnic Institute

Blacksburg, Virginia

June, 1962

ED003044

34144
143

PREFACE

Research in education and meteorology is frequently characterized by an abundance of records, and the experimenter is confronted with the task of ordering them in terms of importance or selecting the relevant ones. The studies contained in the present report illustrate some proposed statistical methods which may be helpful in making a choice among variables (tests, behavior ratings, etc.).

In the search for appropriate methods for the analysis of a given problem there is considerable temptation of professional particularism. Many methods of educational, biological, agricultural, or engineering statistics differ merely in notation and semantics. Frequently, models and methods of analysis have been developed from first principles in one branch of science, even though all essentials were available in another. Algebra is such a powerful tool that a few hours spent in algebraic manipulations may produce a great number of new or modified methods of analysis.

We attempted to apply only standard methods in the present situation. The emphasis was on extensive tryout of each method recommended, and many modifications were made when applications did not lead to satisfactory results. The attempt to identify groups of variables, influenced by some common characteristics, led naturally to the suggestion of factor analysis as a tool. However, the determination of the appropriate number of factors and the study of preferred structures was so crucial that mathematical formulations had to be submitted to many tests before they could be reasonably adopted. A set of programs was developed and tried out (Appendices B and C) in order to establish the usefulness of test criteria, approximations, and preferred structures. The results, which are augmented by additional examples in Chapters IV and V, were very convincing, and the consistency of correct interpretations of known structures should be strong evidence in favor of factor analysis as a tool for finding efficient classification principles.

Validation of educational measurement cannot always be done in terms of scores on some final test; frequently one can merely decide that a person was successful or unsuccessful, or that a person should have been assigned to a particular group, career, or track. It was thus natural to look for a statistical technique of validation which includes all these types of end results as special cases, and a broad interpretation of discriminant functions seemed useful. The function itself is not new; it has been used, under different names and for different purposes, by many authors. It was considered useful, however, to show in this context the broad spectrum of application of one single mathematical function. Again, before proposing this technique as a tool, we conducted many studies, contained in this report, which showed its usefulness.

The only claim to originality of ideas contained in this report concerns the interpretation of a discriminant function in terms of its relation to observed variables, and the combination of two techniques - factor and discriminant analysis - for the purpose of representative selection and ordering. Several examples throughout the text, especially in Chapters IV and V, illustrate the desirability of selecting variables in terms of the underlying common factors which they represent.

With the advent of high-speed electronic computers, considerations of shortcuts in numerical work seem futile. However, some algebraic manipulation of formulas for the purpose of finding efficient and simple numerical techniques of analysis proved very helpful, even though electronic calculators were available and used. The development and "debugging" of programs for electronic computers takes considerable time, and sometimes an hour spent in algebraic simplification may save days of programming work. Consequently, such algebraic shortcuts have been reported.

One reason for the general scarcity of application of multivariate analysis seems to be an apprehension, on the part of potential users, of the complexity of analytical methods. For this reason it was felt that the inclusion of a section (Chapter II) showing the simplicity of generalization, and a detailed numerical example (Appendix A), might serve a useful purpose. The theorems which prove the correctness of these analogies between univariate and multivariate analysis are quite well known and included in standard textbooks. It is the purpose of the exposition in Chapter II to encourage the user to utilize these analogies. If he is confronted with a complicated design in a multivariate experiment, he should not waste any effort trying to develop the method of analysis from first principles. It is sufficient to refer to a worked-out procedure for the corresponding univariate analysis, which is most likely contained in any standard textbook on experimental design. The generalization to multivariate analysis is trivial.

An attempt has been made to select demonstration studies and educational and anthropological applications which may be typical of many experiments performed in practice. It is hoped that sufficient detail on numerical procedures has been included, so that a potential user may apply the techniques for his own purposes by changing parameters and numbers only. We did, however, make extensive use of matrix formulations and operations, because of the compactness, generality, and superiority of this notation over multiple summations. It must, however, be emphasized that one might easily become a slave of this notation and perform a lengthy matrix operation where a simple summation could have saved time. We tried to avoid this pitfall, and leave it to the reader of this report to show us some examples of unnecessarily cumbersome operations which are quite certainly contained somewhere in the text.

My sincere thanks are extended to Robert H. Brown and Robert A. Bales who pointed out shortcuts, helped in the programming and conducted diligently many of the studies which were needed to support or discard techniques. Some of the work contained in Appendix B was supported, prior to the Cooperative Research contract, by a grant from the McDermott Foundation.

I also wish to express my gratitude to the staff members of the Department of Statistics at the Virginia Polytechnic Institute, for their helpful comments and criticism, and to Mrs. Lela Barnhisel for her excellent work in typing this report.

Blacksburg, Virginia
June, 1962

Rolf E. Bargmann

TABLE OF CONTENTS

Page

PREFACE i

I. INTRODUCTION 1

1. Choice of Experimental Units 1
2. Choice of Response Variables 6
3. Discrimination and Classification 11
4. Notation 13

II. SOME ANALOGUES IN UNIVARIATE AND MULTIVARIATE ANALYSIS 15

1. The General Linear Model in Univariate Analysis 15
2. Formulation and Testing of Hypotheses in Univariate Analysis 22
3. General Linear Model in Multivariate Analysis 26

III. A GENERALIZED DISCRIMINANT FUNCTION 31

1. Non-Centrality Parameters 31
2. Definition and Estimation of a Discriminant Function 35
3. Example - 2 groups 37
4. Example - k groups 39
5. Example - Cross Classification 42
6. Example - Multiple Regression 54
7. Summary 57

Page

IV. REPRESENTATIVE ORDERING OF VARIABLES 59

1. Statement of the Problem 59
2. Description of a Demonstration Study 63
3. Ordinary Discriminatory Analysis 65
4. Factor Analysis Discrimination 73
5. Step-Down Analysis and Interpretation of Results 92
6. Conclusions and Recommendations 94

V. REPRESENTATIVE SELECTION OF VARIABLES 97

1. Statement of the Problem 97
2. Theory of Representative Selection 103
3. An Artificial Demonstration Study 107
4. Example of Representative Selection of Variables in an Educational Experiment . . . 137

REFERENCES 164

APPENDIX A: Multivariate Analysis of a Standard Two-Way Classification

APPENDIX B: A Comparison of the Maximum Determinant Solution in Factor Analysis with Various Approximations

APPENDIX C: IBM 650 Programs for Factor Analysis

I. INTRODUCTION

1. Choice of Experimental Units

The concept of experimental design is perhaps as old as the scientific method itself. While statisticians may make extensive use of this concept, they certainly did not introduce it. Suppose a chemist were to take measurements of pH values and specific gravity using different kinds of acids and bases indiscriminately. To be sure, on the basis of such determinations, he may compute mechanically some indices of association, e.g., the correlation between these two measurements. But such a numerical index is devoid of any meaning, for the experiment was not designed for this purpose. In fact, this somewhat absurd example is an illustration of a totally undesigned experiment.

Even in the educational and social sciences, where strictly controlled laboratory experiments are rarely feasible, some experimental design, or at least some general idea as to the purpose of collecting and analyzing data, should precede the analysis. The first step in such considerations is the definition of groups or "populations" which are to be studied. One must then take representative samples of such groups. Each unit of such a sample will be called an experimental unit, following a suggestion by S. N. Roy. In educational research these units are usually persons.

Groups or populations may be defined in terms of different treatments applied. Such "treatments" could be, e.g., different teachers assigned to a classroom, different textbooks used, different types of schools or curricula, etc. In addition, each experiment must contain one or more responses to the treatments, such as grades attained, scores on achievement tests, or number of dropouts. The governing principle in the selection of experimental units is that of adequate representation. Ideally, such units would be chosen so that the only difference between members of different groups is in terms of the treatment effect under study. However, in educational research it is practically impossible to find groups of people who, for example, are identical in every respect except for assignment to different tracks in a high school. We encounter the almost overriding influences of intelligence and environmental background which may be so strong as to obliterate any secondary effects due to differences in classroom assignment.

Is it impossible, then, to study the effect of such treatments in the presence of much stronger effects which we recognize but do not wish to consider at the moment? Surely, if we could somehow eliminate the stronger effects, we may raise the secondary ones to primary status. In the famed Eight Year Study (see, e.g., [1]) equalizing of extraneous effects was performed by the tedious process of finding a match to every person according to many different criteria. This approach is in many respects ideal in that it allows the experimenter to control physically the effect of each extraneous variable, but it is very costly and time-consuming.

A less satisfactory but simpler approach consists of making measurements on each extraneous variable and eliminating its effect by some mathematical technique. Such variables are termed concomitant variables, as distinguished from response variables which were defined earlier. Statistically they must satisfy certain criteria which will be discussed in a later section. Where concomitant variables are not characterized by continuous measurements but by the fact that the individual belongs to a certain group, not presently under study, their effects can be equalized through the use and elimination of block effects. The score on an intelligence test may be considered as an example of a continuous concomitant variable; the fact whether a college student attended a college-preparatory or general education track in high school may serve as an example of block effects. The disadvantage of this mathematical equalization of extraneous effects versus the matching of individuals lies in the fact that the former requires some mathematical abstraction of the physical situation. Even with the most careful and relevant choice of a mathematical model there will always be the possibility that treatment effect differences may be camouflaged or spuriously intensified by the presence of unrecognized extraneous effects, or by the choice of an inadequate mathematical model in explaining the known ones.

The foregoing considerations pertain to the choice of experimental units to be included in a given study. After definition of all treatment effects, block effects, and concomitant variables one will certainly try to find several individuals or experimental units with identical group characteristics (treatment and block effects) so that one may draw statistical inferences concerning the populations from which such individuals arose. In the statistical treatment of the General Linear Model (see Chapter II) such samples within a given group are assumed to be random samples, i.e., the residual or error effects for different individuals are independent of each other, and positive and negative departures from the aggregate of all effects within a given group are equally likely. If we are willing to assume further that the error distribution is smooth (continuous and twice differentiable), that large deviations are less likely than small ones, and that the maximum-likelihood estimate of the aggregate effect within a group should be the arithmetic mean of all observations in the group, we can deduce that the errors must be normally distributed (see, e.g., [22]). Of course, in choosing groups and experimental units, one will have in mind the kind of response variables one wishes to measure, but it must be noted that, mathematically, the process of choosing groups is distinct from the process of selecting response variables. Once groups have been defined and experimental units selected, the question whether one or several response variables are to be measured, or which variables are to be included, is essentially unanswered. Let us suppose that we have the j'th individual under the second treatment effect (e.g., in an accelerated class) who happens to have the third block effect (e.g., comes from a professional family) and a score of 54 on some intelligence test (which is assumed to be eliminable as a quadratic function). We may then say

$$E(y_{23j}) = \mu + \tau_2 + \beta_3 + 54\gamma_1 + 54^2\gamma_2 ,$$

where E stands for expected value of or expectation of. The parameters, μ (a general effect), τ (treatment effect), β (block effect), γ_1 (linear effect on intelligence test score), and γ_2 (quadratic effect) are assumed to determine adequately the response variable y (which may be the score in some achievement test). If y stands for another achievement test, the numerical values of the parameters will, of course, be different, but the mathematical model stated above will remain the same. Thus, after identification of the groups and concomitant effects (the design) one still has to investigate the criteria for choosing response (y) variables.

2. Choice of Response Variables

In the choice of response variables on which measurements are to be made one is usually led by scientific hypotheses regarding the nature of these variables, possibly quite unconfirmed. A set of variables will be chosen such that its members belong, at least logically, or circumstantially, to one or very few groups with common characteristics. Nothing prevents an experimenter from deliberately including unrelated variables in his set of responses, but the scientific value of any interpretation based upon such a procedure would be questionable. More commonly, the experimenter will take response variables pertaining to a common physical class. A set of response variables may consist of different achievement tests, different aptitude tests, or assessment of personality characteristics. If the experimenter does not shy away from complication and resulting difficulties of interpretation, he may even include members of all three groups. However, he would certainly refrain from including anthropological measurements (except, perhaps, as concomitant variables) or metabolic data. Let us assume, for simplicity, that all response variates belong, logically, to one homogeneous class, for example, they may all pertain to some kind of mental achievement. Even if the experimenter makes a deliberate effort to take response variables from a homogeneous set he may not succeed in selecting them in such a way that they are primarily influenced by one and only one underlying factor. As a matter of fact, he may select the variables and perform measurements on them for the very purpose of deciding whether they have a bearing on some common phenomenon. If he finds that they belong to more than one common group he may wish to select a few representative response variables from each common group and subject these to further study. This is, viewed in a somewhat different light, the problem of Factor Analysis [23].

A mathematical model is useful to the applied scientist only to the extent that it reflects and simulates the reasoning process which he applies to the interpretation of his experiments. To regard the choice of response variables as some process of random sampling from an infinite universe of such variables may be of some mathematical interest; in the majority of scientific applications, however, this model is unrealistic. Inasmuch as any mathematical model entails certain restrictions and idealizations, we will try to state some of the limitations which apply to our approach to the problem of representative selection and ordering of response variables. The experimenter may then decide whether he can be satisfied with this approximation to nature and he will, hopefully, refuse to use these methods if the assumptions are untenable.

1. It is assumed that response variables are selected and measured not for the purpose of individual study of each variable, but for the purpose of studying one or very few common phenomena which affect each of the response variables.

2. Response variables must fall into some group such that the members in each group are related to each other more closely than to members outside the group. (A given response variable may, however, belong to more than one group).
3. Except for the sharing of common group characteristics the response variables will be independent, (Principle of common and unique factors [23], or Principle of conditional independence [3, 4]). This assumption states that each response variable consists of a common part, which it shares with at least one other response variable, and a unique part which, though not limited in relative size, is uncorrelated with the unique part of any other response variable.
4. Any description of groups of experimental units, any discrimination between them or assignment of experimental units to one of them, will be in terms of the underlying common parts of the response variables.

Assumption 4 is probably the most important one for educational research. We may, for example, classify a high school freshman into one of three tracks on the basis of an achievement test in English, or in mathematics, or in science. Each will result in a somewhat different assignment policy. We may take some weighted or unweighted total of the three scores based upon subjective evaluation of the importance of each score, or upon validity studies based upon some success criterion. Each of these methods has some merit. In the methods proposed in this report, however, we will use assumption 4 as the basis of classification, i.e., assign him in terms of his estimated performance on that common part which influences all three achievement scores. Furthermore, if variables are to be eliminated for future assignment purposes (for example, if it would take too long to administer all the tests used in a pilot study) the strongest representatives of each group will be retained.

There seems to be some misunderstanding concerning such common parts or common factors. They are sometimes considered as real physical entities, such as mental or anthropological factors attached to genes. A simple example from chemistry may serve to dispel such ambitious interpretations. The elements Li, Na, K, Rb, Cs, and the radical NH_4, belong to some common group in terms of physical and chemical properties of their compounds. They are, or behave like, alkali metals. These alkali metals have many common properties; even the group itself has some group characteristics in that, e.g., each atom has just one electron in the outer shell. But "the alkali metal" has no physical existence, it is merely a conceptual unit useful and efficient for description of the members belonging to this group. The radical NH_4 is also of considerable interest; it behaves like an element and, if we have to assign it to any group of elements, the alkali group is the logical one.

Nor should it be assumed that chemistry has an easy task compared with educational research. If in the foregoing example the intensity of reaction of the element with water had been used as a response variable, Calcium would have been included in the group though, by other criteria (such as valence) it does not belong. If valence had been used as a response variable many other elements would have been included leading to a class whose group characteristics would be somewhat difficult to describe.

The statistical technique known as Factor Analysis, (or Factor Analysis "with unknown communalities", as some authors prefer to call it who include many different techniques under the one common name) if used properly, can help to identify groups of response variables. Its results are conclusive and useful _if the choice of response variables to be measured was prompted by considerations such as those mentioned in assumptions 1 to 3_. Just as an undesigned experiment may lead to mathematical artifacts which are devoid of any scientific meaning, factor analysis applied to a set of response variables which were not chosen in some systematic manner will lead to factor loadings which are no more than decorative numerical ornaments. The criticism here is, of course, the same as that applying to the indiscriminate use of statistical indices, such as correlations or variance components, where the data were not collected for this purpose and a corresponding mathematical model was not contemplated. This is not to say that a statistician should refuse to analyze data which were not designed by a statistician; he should merely look for the scientific design and purpose of the experiment before applying some technique of analysis.

3. Discrimination and Classification

Let us assume that the number of students to be admitted to some course is to be restricted, and admission is to be based upon performance in an aptitude test consisting of five part scores. How can such a test be validated? What combination of part scores should be used? The most frequent practice in psychological research seems to be that of correlating each part score with the score on some end-of-course test, and choosing weights in such a way that the correlation between criterion (end-of-course test) and the weighted sum of predictors (part scores) is maximized.

A similar task could be performed by the use of a discriminant function. As in the above mentioned example of validation we have results of five part scores. For calibration purposes, we choose only two groups: those students who were unquestionably successful in the course, and those who were clearly unsuccessful, leaving out of consideration any students for whom a clear decision cannot be made one way or the other. Now, retrospectively, one can obtain some statistical indices, such as the mean vector and variance-covariance matrix, of the five part scores in each of the two groups. One can then ask: What is the combination of part scores leading to a weighted total score with maximum separation between the groups? Of course, the maximum must be expressed in some relative measure, for otherwise one might increase the difference merely by multiplying all weights by 10. The statistical problem is as follows: Find a linear combination of the five part scores,

$$z = \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_3 + \alpha_4 y_4 + \alpha_5 y_5$$

such that the standardized difference between the means in the two groups is a maximum. The solution to this problem is quite simple and contained in standard textbooks (see, e.g. [18]). The function z (or sometimes just the vector $\alpha_1, \alpha_2, \ldots \alpha_5$) is called the linear discriminant function. Once the α_i's have been estimated from a "calibration sample" one may readily use this discriminant function for the determination of a weighted score for every subject. We may even obtain probabilities of misclassification in such a case, under certain simplifying assumptions (see, e.g., [2]). The concept of a linear discriminant function can be easily generalized to such cases where more than two groups are involved, following an approach proposed by S. N. Roy [20] for the purpose of constructing test statistics in multivariate analysis. This generalization is discussed in

Chapter III. The generalized discriminant function will be used as an artificial variable on which the selection and ordering of variables is based. The combination of this generalized discriminant analysis with some aspects of factor analysis will then be used for the representative selection and ordering of variables.

4. Notation

(1.4.1) A matrix will be noted by a capital letter, (e.g., A, F, C). If the matrix has parametric elements, a capital Greek letter will be used, (e.g., Σ). Where the capital Greek letter may be confusing, the corresponding Roman letter will be used with a tilde underneath (e.g., $\underset{\sim}{R}$, since capital ρ, i.e., P, would be read as p).

(1.4.2) If the order of a matrix is needed for an understanding, it will be denoted by the number of rows and columns in parentheses. For example, A(n x m) indicates that matrix A has n rows and m columns.

(1.4.3) A column vector will be denoted by an underscored letter, generally lower-case, (e.g., $\underline{a}$, $\underline{\phi}$) sometimes capital letter underscored (e.g., $\underline{T}$ for treatment totals). A row vector will have a prime after the letter (e.g., $\underline{y}'$).

(1.4.4) A determinant will be indicated by vertical bars (e.g., $|R|$, $|R - FF'|$).

(1.4.5) A' = transpose of A.

(1.4.6) D_a denotes a diagonal matrix with elements a_i in the principal diagonal and zeros elsewhere.

(1.4.7) $\underline{\xi}$ = column vector of parameters in the univariate model.

(1.4.8) ϕ = matrix of parameters in the multivariate model.

(1.4.9) A cap ($\hat{}$) above any letter denotes an estimate.

(1.4.10) Σ = variance-covariance matrix

(1.4.11) S = estimate of variance-covariance matrix (the kind of estimate will be indicated in the text).

(1.4.12) $\underset{\sim}{R}$ ("capital rho", see 1.4.1), with elements ρ_{ij}, is the population correlation matrix

(1.4.13) R = sample correlation matrix

(1.4.14) SSH sums-of-squares due to hypothesis

(1.4.15) SSE sums-of-squares due to error

(1.4.16) n_h degrees of freedom due to hypothesis

(1.4.17) n_e degrees of freedom due to error

(1.4.18) E "Error Matrix" = matrix of sums of squares and cross-products due to error (multivariate generalization of SSE)

(1.4.19) H "Hypothesis Matrix" = matrix of sums of squares and cross-products due to hypothesis (multivariate generalization of SSH)

(1.4.20) $\mathrm{var}(\underline{x})$ Matrix of variances and covariances of all x_i and (x_i, x_j) pairs.

II. SOME ANALOGUES IN UNIVARIATE AND MULTIVARIATE ANALYSIS

1. The General Linear Model in Univariate Analysis

In many situations arising in controlled laboratory experimentation, measurements of variables affecting the experiment can be made with great precision so that, theoretically, they may be regarded as knowable without error. Such variables are called concomitant variables. As illustrations for this kind of variables we may regard the time at which an experiment is to be performed, or a controlled temperature. If, e.g., measurements are to be taken for ten minutes, at one minute intervals, and if the temperatures of different runs are to be kept constant at 20 and 40 degrees centigrade, these conditions can be realized to any desired degree of accuracy. Dichotomies can also be regarded as measurements on a concomitant variable, for one may define a variable x_i, with observations x_{ij}, which take the value 1 if the j'th experiment falls into the i'th group, and 0 otherwise. Naturally, it would be peculiar to speak about "standard errors" of such variables x_i which, by the nature of their definition, can take a value 1 and 0 whenever the experimenter wishes them to take that value. For the same reason, the other concomitant variables whose values are determined before the experiment, do not admit of a "standard error" or probabilistic interpretation.

In the majority of physical experiments there is only one variable, usually designated by y, which is not predetermined. This is called the response variable, predictand, or criterion. It is a random variable, i.e., a probability is attached to each value or range of values which this variable can take. In the above illustration, the amount of product formed would be an example of a measurement of this kind. Suppose we find that 20.5 mg of product is formed after 5 minutes, at a temperature of 40 degrees centigrade. If we repeat the experiment under identical conditions we will not expect exactly 20.5 mg the second time. We may, after thorough investigation, decide that 20.5 ± 0.2 mg will be the outcome, with a probability of .99, but, in a given experiment, however carefully conducted, we will not be able to predict the exact outcome. It is in this situation that a statistical approach is utilized for the purpose of making inferences regarding future behavior of the system under study. This analysis, which is based upon one or more concomitant variables and one response variable, is called Univariate Analysis. Simple and multiple regression, analysis of variance and covariance (however complicated the design or model may be) are examples of this technique.

A rather general model has been formulated to represent, in a compact mathematical form, all the situations which fall into this category. It is called the General Linear Model (see, e.g., [9, 13]). It can be formulated as follows: Let N observations of a random variable y be denoted by $y_1, y_2, \ldots y_N$ and written out as one long column vector $\underline{y}$ (N x 1) (read "vector $\underline{y}$ with N elements" or "matrix $\underline{y}$ with N rows and 1 column"). Let there be m possible "effects" or "parameters", $\xi_1, \xi_2, \ldots \xi_m$, which are written as a column vector $\underline{\xi}$. This vector would include main effects, group effects, effects of the i'th treatment at level k, fixed block effects, regression coefficients, coefficients of polynomials of a concomitant variable, etc. The General Linear Model then simply states

(2.1.1) $E(\underline{y})(N \times 1) = A(N \times m)\underline{\xi}(m \times 1)$,

which reads, "Expectation of $\underline{y} = A\underline{\xi}$", where A is a matrix with N rows and m columns called the "model" or "design" matrix. It contains ones and zeros indicating whether an effect is present or absent in a given experimental unit, or the values of one or more continuous concomitant variables. For example, in a simple regression situation with 5 observations, we have the model

$$E(y_i) = \alpha + \beta x_i \ , \quad i = 1, 2, \ldots 5$$

which can be written, in full, as

Example 2.1

$$E\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} .$$

$$E(y) = A \quad \underline{\xi}$$

Note that the values x_i, in this case, are assumed known without error. They are the values of the concomitant variable x. By contrast, the operator E (expectation) is placed in front of the vector $\underline{y}$ indicating that the values of the y_i's are not specified. The parameters, α and β, are the assumed or true effects which one may estimate from a sample of observations on y.

In the case of analysis of variance with three groups, the model is

$$E(y_{ij}) = \ + \alpha_i \quad (i = 1, 2, 3; \ j = 1, 2, \ldots n_i)$$

which means that the j'th observation (i.e., every observation) in group i is affected by the general effect, μ, and the i'th group effect, α_i. In the notation of the General Linear Model, assuming 3 observations in the first group, 4 in the second, and 2 in the third group, this model can be written as:

Example 2.2

$$E\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad .$$

$$E(\underline{y}) \qquad\qquad A \qquad\qquad \underline{\xi}$$

A simple combination of the two previous examples, regression and analysis of variance, is the model of analysis of covariance

$$E(y_{ij}) = \mu + \alpha_i + \beta x_{ij} \quad ,$$

where x_{ij} denotes an observation on a concomitant variable x. For example, in an educational situation, we may have boys and girls, and record the chronological age of each subject, so that α_1 and α_2 may be the effects due to sex and x_{1j} is the age of the j'th boy in the experiment; y_{2j} may be the score on an achievement test for girl No. j. In terms of the general linear model:

Example 2.3

$$E\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & x_{11} \\ 1 & 1 & 0 & x_{12} \\ 1 & 1 & 0 & x_{13} \\ 1 & 0 & 1 & x_{21} \\ 1 & 0 & 1 & x_{22} \\ 1 & 0 & 1 & x_{23} \\ 1 & 0 & 1 & x_{24} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix} .$$

$$E(\underline{y}) \quad = \quad A \quad \underline{\xi}$$

The word "linear" may be somewhat misleading in this context; it merely indicates that parameter combinations be linear. It is not necessary that the combination of concomitant variables be linear. Take, for example, a model which is assumed to be of the form

$$E(y_i) = \alpha + \beta x_i + \gamma x_i^2 + \delta \log x_i \ .$$

This non-linear relationship can nevertheless be put into the "General Linear Model" form as:

Example 2.4

$$E\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \log x_1 \\ 1 & x_2 & x_2^2 & \log x_2 \\ 1 & x_3 & x_3^2 & \log x_3 \\ 1 & x_4 & x_4^2 & \log x_4 \\ 1 & x_5 & x_5^2 & \log x_5 \\ 1 & x_6 & x_6^2 & \log x_6 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} .$$

$$E(\underline{y}) \quad = \quad A \quad \underline{\xi}$$

It is thus seen that even a very complicated "response surface" can be reduced to the General Linear Model form.

The significance of this compact mathematical formulation may be evident even from these few examples. Once we establish results and methods of estimation based upon this general model, we know that they apply to a very comprehensive class of situations. In order that a meaningful analysis of this model may be performed certain assumptions must be made regarding the variability of $\underline{y}$. However, the standard assumptions of independence and equal variance of errors are not at all necessary, they merely reduce the problem of estimation to a very simple form, the so-called "normal equations" for the estimation of $\underline{\xi}$, i.e.,

(2.1.2) $$A'A\,\hat{\underline{\xi}} = A'\underline{y}\ .$$

If the variances and covariances of the observations are all different but specified except for an arbitrary constant as

$$\mathrm{var}(y_i) = k\sigma_{ii}\ , \quad \text{and} \quad \mathrm{cov}(y_i, y_j) = k\sigma_{ij}\ ,$$

the normal equations will be

(2.1.3) $$A'\Sigma^{-1} A\,\hat{\underline{\xi}} = A'\Sigma^{-1}\,\underline{y}\ ,$$

where Σ is a symmetric matrix of order $N \times N$ with elements σ_{ij}, and Σ^{-1} denotes the inverse of this matrix. Naturally, this system of linear equations in $\hat{\underline{\xi}}$ is more difficult to evaluate but, theoretically, best linear estimates can always be obtained regardless of the complexity of the model (see, e.g., [5]).

2. Formulation and Testing of Hypotheses in Univariate Analysis

Aside from the problem of estimating the unknown parameters and thus predicting the values of future observations under similar experimental conditions we may wish to know whether there is a difference between effects, or whether regression is present or absent, or whether a quadratic term contributes or not, etc. In a sample we will always find some differences or departures from a "null hypothesis" but we may wish to ascertain whether such differences are "significant" or whether they could have arisen, by mere chance, with an appreciably large probability. If chance suffices to explain the observed discrepancies we would not look for a more sophisticated interpretation. Only if the probability that the observed discrepancies could have arisen by chance is very small, less than .05 or .01, say, will we decide that the discrepancies are "significant" and proceed to interpret or find reasons for such effect differences. One starts by formulating a null hypothesis, e.g., H_0: $\beta = 0$ in the case of simple regression, or H_0: $\mu_1 = \mu_2 = \mu_3$, in the case of analysis of variance, or even H_0: ($\beta = 0$ and $\mu_1 + \mu_2 = \mu_3 + \mu_4$) in an analysis of covariance. In most instances, these hypotheses are statements of no effect differences, or statements of certain theoretical laws which may or may not be borne out by the experiment. Again we can formulate a wide variety of hypotheses, in a mathematically compact form, as a General Linear Hypothesis:

$$C\,\underline{\xi} = \underline{k} \quad ,$$

where C is a matrix with n_h rows and m columns, $\underline{\xi}$ is the parameter vector of the General Linear Model, and $\underline{k}$ is a column vector with n_h known elements, frequently expressed in such a way that they are all zero. If the hypothesis is stated without redundancies, n_h is called the number of "degrees of freedom due to hypothesis", and plays a central role in the testing of hypotheses.

Let us again consider the examples in section 1 of this chapter. In the regression case (Example 2.1), the hypothesis of absence of regression can be stated as

$$\underset{C}{\begin{bmatrix} 0 & 1 \end{bmatrix}} \underset{\underline{\xi}}{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}} = \underset{0}{0} \quad .$$

Here, the matrix C is a simple row vector with elements 0, 1. In the analysis-of-variance example (Example 2.2), equality of the three group effects, i.e.,

$$H_0\text{:} \;\; \mu_1 = \mu_2 = \mu_3 \quad ,$$

can be stated as

$$\underset{C}{\begin{bmatrix} 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}} \underset{\underline{\xi}}{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}} = \underset{0}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}} .$$

Note that the two rows of this expression read:

$$\alpha_1 - \alpha_3 = 0 \quad \text{and} \quad \alpha_2 - \alpha_3 = 0 \ .$$

Any further statement (e.g., $\alpha_1 - \alpha_2 = 0$) would constitute a redundancy in that it is implied by the above two statements. Two statements are sufficient to describe the null hypothesis, and thus n_h, the number of degrees of freedom due to hypothesis, is equal to 2. In the analysis-of-covariance example (Example 2.3), we may wish to test for equality of the two group effects (no difference between sexes) and, simultaneously, no effect due to age. The General Linear Hypothesis would thus be

$$\underset{C}{\begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}} \underset{\underline{\xi}}{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}} = \underset{\underline{0}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}} .$$

Of course, we can test each of the two statements separately ("break down degrees of freedom") but we would, in general, not expect the test statistics to add to the statistic for the combined hypothesis.

In the case of a normal distribution of errors, and in the special case of independence and equal variance of errors, two expressions are usually required: Sum of squares due to the hypothesis (SSH) and Sum of squares due to error (SSE). These expressions are, formally, equal to

$$SSH = \hat{\underline{\xi}}'C'[C(A'A)^{-1}\,C']^{-1}\,\hat{\underline{\xi}}\ ,$$

and

$$SSE = \underline{y}'\underline{y} - y'A\,\hat{\underline{\xi}}\ ,\ \text{where}$$

a primed symbol denotes a transpose, $\hat{\underline{\xi}}$ is the solution of the normal equations, and $(A'A)^{-1}$ has to be taken with a grain of salt, for A'A may be singular because of the peculiar formulation of the model. Suffice it to say, in this connection, that A'A can always be made non-singular in such a case by the introduction of a relatively arbitrary set of "constraints" (e.g., $\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 0$) into the normal equations. Then, if the null hypothesis is true,

$$(SSH/n_h) \div (SSE/n_e) = F$$

has the F distribution with (n_h, n_e) degrees of freedom, where n_e denotes the "degrees of freedom due to error" and equals, in general, N - r, where r is the rank of the A-matrix ("effective" number of parameters, i.e., (m - number of constraints needed) = r). The expressions SSH and SSE can be readily generalized to situations where the assumptions of independence and equal variances are not met (see [5]). An important corollary of the above testing procedure is that

$$SSE/(SSH + SSE) = \beta$$

has the Incomplete Beta Distributions (lower tail) with parameters $(n_e/2; n_h/2)$ (see [17]). This fact will be needed for the purpose of finding generalizations to multivariate analysis. The quantity $1 - \beta = R^2$ is often called the square of the "multiple correlation" of y on the concomitant variables, because of the formal analogy of arithmetic procedures with those for obtaining multiple correlations. It is recommended that the practice of referring to "correlations" be discouraged when concomitant variables are involved, for a correlation expresses a relationship between random variables. There is, of course, no objection to using the symbol R^2 for $1 - \beta$ and interpreting it as an index of goodness-of-fit of a regression model.

The foregoing discussion of the General Linear Model and General Linear Hypothesis, though sketchy and incomplete, should be sufficient for an understanding of the multivariate extension of these formulations.

3. General Linear Model in Multivariate Analysis

As explained in Sections 1 and 2 of the Introduction, the choice of y (response) variables is distinct from and independent of the choice of the design. In Section II.2 we assumed only one response variable and developed the model and its implications accordingly. The General Linear Model in multivariate analysis is a straightforward extension of the univariate model. Instead of one column vector $\underline{y}$ of N responses on one response variable we have a matrix Y(N x p) of N responses each on p response variables. Accordingly, instead of one column vector $\underline{\xi}$ of parameters, we have a matrix ϕ(m x p) of parameters *), where each of the p column vectors pertains to one of the response variables. The General Linear Model in multivariate analysis simply reads

(2.3.1) $E(Y) = A\,\phi$,

and is a compact statement of the individual univariate models

(2.3.2) $E(\underline{y}_1) = A\,\underline{\phi}_1 \,,\; E(\underline{y}_2) = A\,\underline{\phi}_2 \,,\; \ldots\; E(\underline{y}_p) = A\,\underline{\phi}_p$,

where $\underline{y}_i$ is the i'th column of Y and $\underline{\phi}_i$ is the i'th column of ϕ. Note, in particular, that A is the same for all response variables. (For more detail, see [20]). The very statement of the model indicates that estimation in the multivariate case is a straightforward generalization of estimation in the univariate case, since one merely needs to estimate each of the vectors $\hat{\phi}_1, \hat{\phi}_2$, etc. by standard univariate methods.*)

---

*) The capital Greek Ξ has not been used because of awkward appearance in print.

For the testing of hypotheses the assumptions are somewhat more restrictive than in univariate analysis. We must assume that the variance-covariance matrix for all y's in a row is the same for all rows, and that row vectors are independent. Formally

$$\text{var}(\underline{y}_i') = \Sigma \qquad (2.3.3)$$

$$\text{cov}(\underline{y}_i', \underline{y}_j') = 0 \; ^{**)} \; .$$

On the assumptions (2.3.3) and the assumption of multivariate normal distribution of "errors", hypothesis testing is a very straightforward analogue to univariate analysis. The normal equations for estimating parameters are simply

$$A'A \, \hat{\phi} = A'Y \quad , \qquad (2.3.4)$$

and a "hypothesis matrix" for the hypothesis

$$C \, \phi = 0 \qquad (2.3.5)$$

---

*) This simple analogy holds, however, only in the case that either var($\underline{y}'$) is the same for each row vector of Y or that var($\underline{y}_i'$) is known, except for a constant, for each row vector $\underline{y}_i$. If some (but not all) observations in a row vector are missing, the analysis becomes quite complicated (see [16]). On the other hand, if all elements of some of the row vectors in Y are missing, the analysis is again simple, for it is merely an extension of univariate analysis with missing values (see [5]).

**) Sometimes a simple linear transformation may produce this result, see example of two-way classification in Chapter III.

can be obtained formally as

(2.3.6) $$H = \hat{\Phi}'C'[C(A'A)^{-1}C']^{-1}C\hat{\Phi}$$

and an "error matrix" as

(2.3.7) $$E = Y'Y - Y'A\hat{\Phi} \quad .$$

These matrices play the same role as SSH and SSE in the Beta statistic (not F) in the univariate case. The diagonal elements of H are simply the SSH expressions for each response variable, the off-diagonals are obtained readily as corresponding cross-product expressions between pairs of variables. Appendix A shows this simple generalization for the example of a two-way classification. It also shows some of the test procedures based upon E and H. The methodological consequence of this formal development is the following simple fact: *If one knows a procedure for obtaining sums of squares for hypothesis* (treatment; treatment, adjusted; interaction; etc.) *and sums of squares for error, one can readily obtain, by straightforward generalization, the matrices H and E needed for multivariate hypothesis testing and confidence region estimation*. The reader who is familiar with computational techniques in the (univariate) analysis of covariance will recognize the cross-product expressions. This does not mean, however, that use of the same numerical techniques in the two cases (univariate covariance analysis and multivariate analysis of variance) makes them mathematically or statistically identical. It should, however, take some of the unjustified scare out of multivariate analysis, which is mostly due to the complexity of distributions involved for testing hypotheses. The latter, however, are well tabulated or readily approximated, with great accuracy by χ^2-series, and the numerical methods of obtaining test statistics or confidence region estimates, tedious though they may be, are simple generalizations of familiar univariate techniques. For the likelihood-ratio test, $|E|/H + E|$ is the generalization of the Beta statistic SSE/(SSH + SSE) in univariate analysis (see Section II.2). Care should be taken not to generalize the F statistic, for $|H|$ is quite frequently equal to zero and useless for statistical analysis. The largest characteristic root of $E^{-1} H$ (which is never obtained directly, but always by some simplifying numerical steps, see Appendix A) is S. N. Roy's [20] "union-intersection" statistic. The associated eigenvector will play a considerable role in the development of a generalized discriminant function presented in the next chapter.

III. A GENERALIZED DISCRIMINANT FUNCTION

1. Non-Centrality Parameters

We assume a general linear model in univariate analysis

(3.1.1) $E(\underline{y}) = A\,\underline{\xi}$

which establishes some mathematical law by which a response variable is related to concomitant variables or group effects. Chapter II contained a discussion of a general linear hypothesis

(3.1.2) $C\,\underline{\xi} = \underline{0}$.

We will now consider the alternative to this hypothesis, namely,

(3.1.3) $C\,\underline{\xi} = \underline{\eta} \neq \underline{0}$.

If one wants to study the power associated with the test of (3.1.2) one has to obtain the parameters entering into the non-central distribution. It is quite easy to show (see, e.g., [18, 20]) that the test statistic

(3.1.4) $F = (SSH/n_h) \div (SSE/n_e)$

has, under the alternative (3.1.3), a non-central F distribution with "non-centrality parameter"

(3.1.5) $\gamma^2 = \underline{\eta}'[C(A'A)^{-1}C']^{-1}\,\underline{\eta}/\sigma^2$, i.e.,

a quantity which is obtained from SSH by replacing estimates of $\underline{\xi}$ by the postulated values of these parameters under the alternative and dividing by $\mathrm{var}(y_i)$.

This non-centrality parameter plays the role of (the square of) a standardized distance function, invariant under change of origin and unit of measurement. Let us assume just two groups of sizes n_1 and n_2. Then SSH = sum of squares between groups $= (\bar{y}_1 - \bar{y}_2)^2/(1/n_1 + 1/n_2)$, where $\bar{y}_1$ and $\bar{y}_2$ denote the means in the two groups (simply the t^2 statistic without s^2 in the denominator). This enables us to write the non-centrality parameter

$$(3.1.6) \qquad \gamma^2 = (\mu_1 - \mu_2)^2 / \sigma^2(1/n_1 + 1/n_2)$$

which is readily recognized as the square of the standardized difference between two means. It should be noted here that the sample sizes enter into this non-centrality parameter for obvious reasons: A given difference, $(\mu_1 - \mu_2)$, would result in a greater value of the power function if the sample were large than if it were small.

As another illustration, let us take the case of a one-way classification analysis of variance, assuming equal group sizes for simplicity. Let each group have n observations. Then

$$SSH = n \sum_{i=1}^{k} (\bar{y}_i - \bar{\bar{y}})^2$$

where $\bar{y}_i$ is the mean of the observations in the i'th group and $\bar{\bar{y}}$ is the grand mean. Consequently

$$(3.1.7) \qquad \gamma^2 = n \sum_{i=1}^{k} (\mu_i - \mu)^2/\sigma^2 \quad .$$

This means that the standardized distance of each group mean from the grand mean is evaluated $((\mu_i - \mu)\sqrt{n}/\sigma)$, squared, and the squares added. It is these dimension-free standardized indices which, like the correlation coefficient for instance, allow comparisons and study of relationships even when units of measurement are different. Before the introduction of such indices, such problems as comparing heights and weights of persons were quite ambiguous. The statistical comparison, regardless of units of measurement employed, simply asks the question: "How many standard units (σ = unit of measurement) is the height of a person above or below the mean height, and how many standard units is the person's weight above or below the mean weight of the group?" Each "standard score" indicates a dimension-free distance. The non-centrality parameters are, in the same sense, statements of a standardized distance (squared) between groups, if under the null hypothesis they are assumed to have equal location parameters.

We will propose the use of this non-centrality parameter as a single index of discrepancy between groups. Just as $C\underline{\mu} = \underline{0}$ indicates a null hypothesis, the quantity $C\underline{\mu}$ induces a grouping into a model.

Take a one-way classification which can be (though usually is not) stated in the non-singular model

(3.1.8) $$E(y_{ij}) = \mu_i$$

or

(3.1.9) $$E(\underline{y}) = \begin{array}{c} (n_1) \\ (n_2) \\ \cdot \\ (n_1) \end{array} \begin{bmatrix} \underline{1} & \underline{0} & \cdots & \underline{0} \\ \underline{0} & \underline{1} & \cdots & \underline{0} \\ \cdot & \cdot & \cdots & \cdot \\ \underline{0} & \underline{0} & \cdots & \underline{1} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \mu_k \end{bmatrix}$$ *)

where $\underline{1}$ denotes a column vector of ones and $\underline{0}$ denotes a column vector of zeros. Now, if C is taken as the identity matrix, $C\underline{\mu}$ simply indicates that there are k groups with means $\mu_1, \mu_2, \ldots \mu_k$ **). With a suitable definition of C, the non-centrality parameter can be used as an overall index of descrepancy between groups.

---

*) The numbers in parentheses in front of the matrix indicate the number of elements in each column vector.

**) Care must be taken in defining a non-centrality parameter here, for H_0: $C\underline{\mu} = \underline{0}$ here would mean that all group means equal zero. If the null hypothesis is to be "all group means equal", the C-matrix must be stated as in Example 2.2, and γ^2 is equal to (3.1.7).

2. Definition and Estimation of A Discriminant Function

We now take a multivariate model

(3.2.1) $E(Y) = A\,\Phi$

and assume that division into groups is indicated by a function $C\Phi$. We wish to find a linear combination of the response variables

$$z = \alpha_1 y_1 + \alpha_2 y_2 + \dots + \alpha_p y_p$$

or

(3.2.2) $\underline{z} = Y\underline{a}$,

such that the non-centrality parameter based upon the variable z and the grouping C is a maximum. Let us take the simplifying assumption $\text{var}(\underline{y}_i') = \Sigma$, the same for all row-vectors and let pairs of row vectors be uncorrelated. Then

(3.2.3) $E(\underline{z}) = A\Phi\underline{a}$

(3.2.4) $\text{var}(\underline{z}) = (\underline{a}'\Sigma\,\underline{a})I$ *)

where I is an (N x N) identity matrix. Then, in terms of $\underline{z}$, the non-centrality parameter is

$$\gamma^2 = \underline{a}'\Phi'C'[C(A'A)^{-1}C']^{-1}C\Phi\underline{a}/(\underline{a}'\Sigma\underline{a})$$

(3.2.5) $\gamma^2 = (\underline{a}'\underset{\sim}{H}\,\underline{a})/(\underline{a}'\Sigma\underline{a})$,

where $\underset{\sim}{H}$ is the parametric analogue of the "hypothesis matrix" H, i.e., a matrix which can be obtained by replacing parameter estimates in H by the postulated values of such parameters under a given alternative.

---

*) For statistical operations with matrices see [5,9]. Note that $\underline{a}'\Sigma\underline{a}$ is a scalar and takes the place of σ^2 in the simple ("least squares") univariate model.

It is very simple to find the maximum of (3.2.5) using techniques of matrix differentiation (see, e.g., [5]), then

(3.2.6) $$\partial\gamma^2/\partial\underline{\alpha} = (2\underset{\sim}{H}\,\underline{\alpha}(\underline{\alpha}'\Sigma\underline{\alpha}) - 2\Sigma\underline{\alpha}(\underline{\alpha}'\underset{\sim}{H}\underline{\alpha}))/(\underline{\alpha}'\Sigma\underline{\alpha})^2 \quad .$$

Equating to zero, replacing $\underline{\alpha}$ by $\hat{\underline{\alpha}}$ and denoting the maximum of γ^2 by $\hat{\gamma}^2$ we have

(3.2.7) $$\underset{\sim}{H}\,\hat{\underline{\alpha}} - \hat{\gamma}^2\,\Sigma\,\hat{\underline{\alpha}} = \underline{0}$$

or

(3.2.8) $$(\Sigma^{-1}\underset{\sim}{H} - \hat{\gamma}^2\, I)\hat{\underline{\alpha}} = \underline{0} \quad .$$

This shows that the largest value which γ^2 can attain is the largest characteristic root of $\Sigma^{-1}\underset{\sim}{H}$, and the maximizing linear combination is $\hat{\underline{\alpha}}'\underline{y}$ (for each row vector of Y), where $\hat{\underline{\alpha}}$ is the eigenvector associated with the largest characteristic root of $\Sigma^{-1}\underset{\sim}{H}$. The elements of $\hat{\underline{\alpha}}$ are relative quantities only, i.e., equation (3.2.8) is still satisfied if each element of $\hat{\underline{\alpha}}$ is multiplied by the same constant.

The quantity $\hat{\underline{\alpha}}$ can be evaluated directly if Σ and $\underset{\sim}{H}$ are defined by some a priori mathematical argument. This situation is rarely encountered in any branch of science, and probably never occurs in educational research.

Rather, one has at hand a "calibration sample", i.e., a sample of observations from different groups based upon experimental units which are "ideal" representatives of each group. The "successful" and "unsuccessful" students mentioned in the example in Section I.3 are examples of such members of a calibration sample. The calibration sample may also be a set of observations as stated in Appendix A, and different group classifications may be induced by different C matrices. From such a calibration sample, the discriminant function can be estimated. Since $\hat{\Phi}$ is a maximum-likelihood estimate of Φ, since the maximum-likelihood estimate of Σ is NE (see Appendix A), since $\hat{\underline{a}}$ can be multiplied by an arbitrary scalar, and because of the invariance property of maximum-likelihood estimates,

$$E^{-1} H \underline{a} = \lambda \underline{a} \quad , \tag{3.2.9}$$

where λ denotes the largest characteristic root of $E^{-1} H$, $\underline{a}$ is the associated eigenvector and, consequently, the maximum-likelihood estimate for the (coefficients of the) linear discriminant function.

3. Example - 2 groups

Let a calibration sample consist of n_1 experimental units with sample mean vector $\underline{\bar{y}}_1' = [\bar{y}_{11}, \bar{y}_{12}, \cdots \bar{y}_{1p}]$ (i.e., means in the first group on each of the p response variables), and n_2 experimental units with sample mean vector $\underline{\bar{y}}_2'$. For each response variable,

$SSH = [n_1 n_2/(n_1 + n_2)](\bar{y}_{1i} - \bar{y}_{2i})^2$, hence H is simply the matrix

(3.3.1) $$H = [n_1 n_2/(n_1 + n_2)](\underline{\bar{y}}_1 - \underline{\bar{y}}_2)(\underline{\bar{y}}_1' - \underline{\bar{y}}_2') \quad .$$

The error sum-of-squares for each response variable is the "pooled" or "within" sum-of-squares

(3.3.2) $$w_i = \sum_{j=1}^{n_1} y^2{}_{1ij} + \sum_{j=1}^{n_2} y^2_{2ij} - n_1\bar{y}^2_{1i} - n_2\bar{y}^2_{2i} \; ,$$

hence, sum of cross-products between variables i and k are

(3.3.4) $$w_{ik} = \sum_{j=1}^{n_1} y_{1ij}y_{1kj} + \sum_{j=1}^{n_2} y_{2ij}y_{2kj} - n_1\bar{y}_{1i}\bar{y}1k$$
$$- n_2\bar{y}_{2i}\bar{y}_{2k} \quad ,$$

which can be combined into an "error matrix" or "within matrix"

(3.3.4) $$E = W \quad .$$

Hence

(3.3.5) $$E^{-1}H = [n_1 n_2/(n_1 + n_2)]W^{-1}(\underline{\bar{y}}_1 - \underline{\bar{y}}_2)\;(\underline{\bar{y}}_1' - \underline{\bar{y}}_2') \, .$$

The eigenvecotrs of a matrix are changed only by a scalar factor if the matrix is multiplied by a scalar, hence the discriminant function is obtained as

(3.3.6) $$W^{-1}(\underline{\bar{y}}_1 - \underline{\bar{y}}_2)(\underline{\bar{y}}_1' - \underline{\bar{y}}_2')\underline{a} = \lambda\, \underline{a} \quad ,$$

where λ is the largest root of the matrix on the left. Now since, except for zero roots, characteristic roots of AB equal characteristic roots of BA (see, e.g., [20]), λ equals the largest root of

$$(3.3.7) \qquad (\bar{\underline{y}}_1' - \bar{\underline{y}}_2')W^{-1}(\bar{\underline{y}}_1 - \bar{\underline{y}}_2) \quad .$$

But (3.3.7) is simply a scalar so that the root is equal to it, hence, substituting (3.3.7) into (3.3.6) one obtains

$$(3.3.8) \qquad W^{-1}(\bar{\underline{y}}_1 - \bar{\underline{y}}_2)(\bar{\underline{y}}_1' - \bar{\underline{y}}_2')\underline{a} = (\bar{\underline{y}}_1' - \bar{\underline{y}}_2')W^{-1}(\bar{\underline{y}}_1 - \bar{\underline{y}}_2)\underline{a} \quad .$$

This is satisfied for $W^{-1}(\bar{\underline{y}}_1 - \bar{\underline{y}}_2) = k\underline{a}$, where k is an arbitrary constant, for after this substitution, (3.3.8) becomes

$$(3.3.9) \qquad k\underline{a}[(\bar{\underline{y}}_1' - \bar{\underline{y}}_2')\underline{a}] = k[(\bar{\underline{y}}_1' - \bar{\underline{y}}_2')\underline{a}]\underline{a} \quad ,$$

where scalars have been set off in square brackets and are seen to cancel. Thus the result is obtained (k is arbitrary)

$$(3.3.10) \qquad \underline{a} = W^{-1}(\bar{\underline{y}}_1 - \bar{\underline{y}}_2) \quad ,$$

which is the familiar discriminant function for two groups (see, e.g., [2,18]) .

4. Example - k groups.

S. N. Roy [20] considered the model

$$E(Y) = A\Phi \quad ,$$

$$\mathrm{var}(\underline{y}_i') = \Sigma \quad ,$$

$$\mathrm{cov}(\underline{y}_i', \underline{y}_j') = 0 \quad ,$$

and the hypothesis

$$C\Theta = 0 \quad ,$$

and asked the question: Which linear combination of the response variables

$$Y\underline{a} = \underline{z}$$

will produce a maximum F-statistic for testing the univariate hypothesis

$C\Theta\underline{a} = \underline{0}$? He proposed the use of this maximum F as a statistic for testing the multivariate hypothesis $C\Theta = 0$. He reasons that, if even this largest possible F statistic, under any linear combinations of the responses, leads to acceptance of the univariate hypothesis, the multivariate hypothesis should be accepted. If, on the other hand, there exists some linear combination such that the univariate hypothesis can be rejected, then the multivariate hypothesis should be rejected ("Union-Intersection Principle"). In our notation, the maximum F turns out to be the largest characteristic root of $E^{-1}H$ and $\underline{a}$ the associated eigenvector.

Thus, if in the k group case, we obtain $\underline{a}$ in this manner, and for each individual in the calibration sample we obtain a composite score

(3.4.1) $$z_i = a_1y_{1i} + a_2y_{2i} + \dots + a_py_{pi} \quad ,$$

where y_{ji} denotes the score of the i'th individual on the j'th response variable, we may perform a univariate analysis of variance based upon the composite z-scores. The F-ratio for the test of equality of group means, based upon this composite, will be as large as it can be for any linear composite score. Thus the composite score

$$z_i = \underline{a}'X_i \tag{3.4.2}$$

with

$$E^{-1}H\underline{a} = \lambda\underline{a} \quad , \qquad {}^{*)}$$

where λ is the largest root, will "dramatize" the group differences. Clearly, such a composite score would be a useful device for classifying future individuals to one of the k groups (i.e., to that group whose mean composite (z) score is closest to the individual's composite score). Also, if some of the response variables (tests, for instance) are to be eliminated in future experiments designed to determine assignment to such groups, one would obviously retain those variables which, in some sense, are most closely related to the composite variable. This will be the basis for selection and ordering of variables as presented in Chapters IV and V.

---

*) Here, $W^{-1}B\underline{a} = \lambda\underline{a}$, where W is the matrix of sums-of-squares and products "within" and B the matrix of sums-of-squares and products "between" groups.

5. Example - Cross-Classification

One may use the argument presented in Section III.4 to indicate the meaning of the discriminant function. However, an approach of reduction to one-way classification will be used here to illustrate the use of a generalized discriminant function.

To clarify concepts, let us consider the following situation: Incoming draftees are to be assigned to one of three groups:

Leadership training (A_1);
Technical duties (A_2);
General duties (A_3);

on the basis of a battery of tests consisting of 10 parts. The educational background influences, of course, performance on these tests. We assume 4 levels

High School not completed (B_1)
High School completed (B_2)
Some college (B_3)
College graduates (B_4) .

A group of, say, 2000 incoming candidates takes the battery of tests and individuals are assigned, by some standard set of criteria, to one of the three A-groups. At the end of one year, definitely successful and properly placed candidates in each of the three A-groups are determined. Suppose there are 1000 of them, and they will make up the calibration sample which may look, in size, somewhat as follows:

Table 3.5.1

	B_1	B_2	B_3	B_4	
A_1	10	90	60	40	200
A_2	40	110	100	50	300
A_3	150	300	40	10	500
	200	500	200	100	1000

The records of these, properly placed, candidates on the test battery can now be studied and subjected to a multivariate analysis of variance with two-way classification, in a manner analogous to that described in Appendix A. The analysis here will be easier in that interactions are not assumed, but more difficult in that the design is not orthogonal, so that A-effects, adjusted for B, must be considered.

To use the unadjusted scores on the ten tests, in some combination, as the basis for assignment, would seem inefficient, since members of group B_4 (college graduates) may tend to attain higher scores than members of group B_1 (unfinished high school) and yet not be suitable for group A_1 (leadership training).

Obviously, a discriminant function based on A-effects is desirable. We will utilise the $E^{-1}H$ approach and study its implications.

To obtain a "sum-of-squares" due to the hypothesis of equal A effects, one proceeds as follows:

a) Let A_1 be a matrix with n rows (here 1000) and r columns (here 3) in which the (i,j) element (element in i'th row and j'th column) is 1 if individual i belongs to group A_j, and zero otherwise.

b) Let A_2 be a matrix with n rows (here 1000) and s columns (here 4) in which the (i, j) element is 1 if individual i belongs to group B_j, and zero otherwise.

c) Let $\underline{j}$ denote a column vector with 1000 elements, all equal to 1.

d) Let $\underline{\mu}'$ denote a row vector with p (here 10) elements indicating "general effects" on each of the ten test scores.

e) Let $\underline{\alpha}_1'$ denote a row vector with p (here 10 elements indicating "specific effect of group A_1" on each of the ten test scores. $\underline{\alpha}_2'$ and $\underline{\alpha}_3'$ are defined accordingly.

f) Let $\underline{\beta}_1'$ denote a row vector with p (here 10) elements indicating "specific effect of group B_1" on each of the ten test scores. $\underline{a}_2'$, $\underline{a}_3'$, $\underline{a}_4'$ are defined accordingly. Then the model is

$$(3.5.1) \quad E(Y) = [\underline{1}, A_1, A_2] \begin{bmatrix} \mu' \\ \underline{a}_1' \\ \underline{a}_2' \\ \underline{a}_3' \\ \underline{a}_1' \\ \underline{a}_2' \\ \underline{a}_3' \\ \underline{a}_4' \end{bmatrix},$$

where the i'th row vectors in Y (score of i'th individual on each of the ten tests) must correspond with the i'th row vector of $[\underline{1}, A_1, A_2]$. This is in the form of the general linear model in multivariate analysis.

g) Let N denote the "incidence matrix", i.e., a matrix whose elements are the sizes of the subsamples. In our example, N is simply the body of Table 3.5.1, a (3 x 4) matrix. (In terms of a) and b), $N = A_1' A_2$).

h) Let $D_{n_{i.}}$ denote a diagonal matrix whose principal elements are the row totals of N. In our example,

$$D_{n_{i.}} = \begin{bmatrix} 200 & 0 & 0 \\ 0 & 300 & 0 \\ 0 & 0 & 500 \end{bmatrix} .$$

(In terms of a) and b), $D_{n_{i.}} = A_1' A_1$).

i) Let $D_{n_{.j}}$ denote a diagonal matrix whose principal elements are the column totals of N. In our example,

$$D_{n_{.j}} \begin{bmatrix} 200 & 0 & 0 & 0 \\ 0 & 300 & 0 & 0 \\ 0 & 0 & 200 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix} .$$

(In terms of a) and b), $D_{n_{.j}} = A_2' A_2$).

j) Let T_a denote a matrix of order $(r \times p)$, here (3×10). The element in the i'th row and j'th column of this matrix is the total (sum) of all (y) observations in group A_i on test j. (In terms of a), $T_a = A_1' Y$).

k) Let T_b denote a matrix of order $(s \times p)$, here (4×10). The element in the i'th row and j'th column of this matrix is the total (sum) of all (y) observations in group B_i on test j. (In terms of b), $T_b = A_2' Y$).

Then the α-effects can be estimated from the well-known (see [5, 9]) adjusted normal equations

$$(3.5.2) \qquad P \begin{bmatrix} \hat{\underline{\alpha}}_1' \\ \hat{\underline{\alpha}}_2' \\ \hat{\underline{\alpha}}_3' \end{bmatrix} = Q ,$$

where $P = D_{n_{i.}} - ND^{-1}_{n_{.j}}N'$

and

$$Q = T_a - ND^{-1}_{n_{.j}}T_b \quad .$$

The matrix P is singular and can be made non-singular by the introduction of an arbitrary constraint, e.g., $\hat{a}'_1 + \hat{a}'_2 + \hat{a}'_3 = 0'$, which amounts to adding a constant to each element of P.

The matrix H (generalisation of sum-of-squares of A effect, adjusted for B effect) is then

$$(3.5.3) \qquad H = Q' \begin{bmatrix} \hat{a}'_1 \\ \hat{a}'_2 \\ \hat{a}'_3 \end{bmatrix} \text{. It is of order } (10 \times 10).$$

The elements of the matrix E are obtained in two steps: (1) obtain a matrix B denoting unadjusted B effects. This is the generalisation of the unadjusted sum-of-squares "between B" effects. Its elements are obtained in the same manner as the rows A or B in Table 5 of Appendix A (where "adjustment" was not required, since the design was orthogonal). Also, the elements of a matrix T (Total) are obtained in exactly the same manner as indicated in Table 5 of Appendix A. Then

(3.5.4) $E = T - H - B$.

(Note that interactions do not occur in the present model).

We will now introduce the notation

$$(3.5.5)\qquad \hat{\Phi}_\alpha = \begin{bmatrix} \hat{\underline{a}}_1' \\ \hat{\underline{a}}_2' \\ \hat{\underline{a}}_3' \end{bmatrix}$$

for the (r x p) (here (3 x 10)) matrix of estimates of A-effects. From (3.5.2) and (3.5.3) we then have

$$(3.5.6)\qquad H = \hat{\Phi}_\alpha' \, P \, \hat{\Phi}_\alpha \quad ,$$

since P is symmetric. Now

$$(3.5.7)\qquad E^{-1} H = E^{-1} \hat{\Phi}_\alpha' \, P \, \hat{\Phi}_\alpha \quad ,$$

and the discriminant function satisfies the relation

$$(3.5.8)\qquad E^{-1} \hat{\Phi}_\alpha' \, P \, \hat{\Phi}_\alpha \, \underline{a} = \lambda \, \underline{a} \quad .$$

Since non-zero roots of AB equal non-zero roots of BA,

$$(3.5.9)\qquad \hat{\Phi}_\alpha \, E^{-1} \hat{\Phi}_\alpha' \, P \, \underline{y} = \lambda \, \underline{y} \quad ,$$

where $\underline{y}$ is the eigenvector associated with the largest root of the (r x r) matrix (here (3 x 3))in (3.5.9). It is, of course, considerably easier to obtain root and vector of this small matrix than of the (10 x 10) matrix $E^{-1} H$.

Premultiplying (3.5.9) by $E^{-1} \hat{\Phi}_\alpha' P$ one obtains

$$(3.5.10)\qquad E^{-1} \hat{\Phi}_\alpha' \, P \, \hat{\Phi}_\alpha (E^{-1} \hat{\Phi}_\alpha' \, P\underline{y}) = \lambda (E^{-1} \hat{\Phi}_\alpha' \, P\underline{y}) \quad .$$

Comparison with (3.5.8) shows that

$(3.5.11)\qquad \underline{a} = E^{-1} \hat{\Phi}_\alpha' \, P\underline{y}$, and thus the problem of finding the largest eigenvector of a 10 x 10 matrix has been reduced to that of finding the vector of a 3 x 3 matrix and performing matrix multiplication, a considerably easier task. An illustration of this will be given in Chapter IV and in Appendix A.

Now we will look at this problem from a different viewpoint. We wish to eliminate B-effects and base the discriminant function on some "purified" set of scores. We can obtain an estimate of general and B effects by taking the means of all observations in groups B_1, B_2, B_3, B_4. Thus, using notation i) and k) of this section,

$$F_b = D^{-1}_{n_{.j}} T_b \quad , \tag{3.5.12}$$

(where the symbol F_b rather than $\hat{\Phi}_\beta$ was used to indicate that these are not the best estimates of B-effects, but merely the unadjusted means.)

The model (3.5.1) can be written as

$$E(Y) = \underline{j}\,\underline{\mu}' + A_1\Phi_\alpha + A_2\Phi_\beta \quad , \tag{3.5.13}$$

where $\Phi_\alpha = \begin{bmatrix} \underline{\alpha}'_1 \\ \underline{\alpha}'_2 \\ \underline{\alpha}'_3 \end{bmatrix}$, and $\Phi_\beta = \begin{bmatrix} \underline{\beta}'_1 \\ \underline{\beta}'_2 \\ \underline{\beta}'_3 \\ \underline{\beta}'_4 \end{bmatrix}$

(3.5.14) Now, let $Y^* = Y - A_2F_b$ denote the matrix of "purified" scores, i.e., scores from which B-effect has been eliminated. Then, since $T_b = A_2'Y$,

(3.5.15) $$Y^* = Y - A_2D^{-1}_{n_{\cdot j}} A_2'Y = (I - A_2D^{-1}_{n_{\cdot j}} A_2')Y \quad .$$

Since $D_{n_{\cdot j}} = A_2'A_2$, the coefficient matrix is idempotent and of rank (n - s) (here 996).

Now, from (3.5.15) and (3.5.13)

(3.5.16) $$E(Y^*) = (I - A_2D^{-1}_{n_{\cdot j}} A_2') \underline{j}\, \underline{\mu}' + (A_1 - A_2 D^{-1}_{n_{\cdot j}} A_2'A_1)\phi_\alpha + (A_2 - A_2D^{-1}_{n_{\cdot j}} A_2'A_2)\phi_\beta \quad .$$

The column vector $A_2'\underline{j}$ contains elements $n_{\cdot j}$, hence $D^{-1}_{n_{\cdot j}} A_2'\underline{j}$ is simply $\underline{j}$ with s (here 4) elements. Each row of A_2 has one and only one unit element, hence $A_2\underline{j} = \underline{j}$, and thus the first term in (3.5.16) cancels. The last term cancels since $D_{n_{\cdot j}} = A_2'A_2$, hence

(3.5.17) $$E(Y^*) = (I - A_2 D^{-1}_{n_{\cdot j}} A_2')A_1\phi_\alpha \quad ,$$

and hence, as desired, the observation vectors in Y* depend only upon parameters α. Since, by (3.5.15), $Y^* = (I - A_2 D^{-1}_{n_{\cdot j}} A_2')Y$, the usual "least squares" assumptions are not satisfied. Now, since $(I - A_2 D^{-1}_{n_{\cdot j}} A_2')$ is idempotent, of order n x n, and rank $(\overline{n - s} \times \overline{n - s})$, one can always find some matrix L $(n \times \overline{n-s})$ such that

(3.5.18) $LL' = (I - A_2 D^{-1}_{n_{\cdot j}} A_2')$,

and a matrix L^* ("inverse from the left"), such that $L^*L = I(\overline{n - s})$, where $I(\overline{n - s})$ denotes the identity matrix of order $n - s$.

We now introduce

(3.5.19) $Z(\overline{n - s \times m}) = L^*(\overline{n - s \times n})Y^*(n \times m)$, where

$m = r + s + 1$. Selecting the k'th column vector of Z one finds

(3.5.20) $\underline{z}_k = L^* \underline{y}_k^*$.

However, in terms of the k'th column vector of Y,

(3.5.21) $\underline{y}_k^* = (I - A_2 D^{-1}_{n_{\cdot j}} A_2')\underline{y}_k$, by (3.5.15)

$= LL' \underline{y}_k$, by (3.5.18) .

Hence, by (3.5.20),

(3.5.22) $\underline{z}_k = L^*LL' \underline{y}_k = L' \underline{y}_k$.

Since $\mathrm{var}(\underline{y}_k) = \sigma_{kk} I$, where σ_{kk} is the k'th diagonal element of Σ,

(3.5.23) $\mathrm{var}(\underline{z}_k) = \sigma_{kk} L'L$.

Because of the fact that LL' is idempotent, i.e., $LL'LL' = LL'$, it is seen that L' is a possible choice for L^*. Hence $L'L = I$, and

(3.5.24) $\mathrm{var}(\underline{z}_k) = \sigma_{kk} I$.

In the same manner it can be shown that

(3.5.25) $\mathrm{cov}(\underline{z}_k, \underline{z}_\ell) = \sigma_{k\ell} I$.

Hence, the z's follow standard least-squares assumptions. Using the fact that L' is a possible choice for L^*, and combining (3.5.19) and (3.5.17), we obtain

(3.5.26) $E(Z) = L'LL' A_1 \Phi_\alpha = L' A_1 \Phi_\alpha$.

By (3.5.24) and (3.5.25), we find

(3.5.27) $\mathrm{var}(\underline{z}_1') = \Sigma$,

so that standard least-squares normal equations can be used for the estimation of ϕ_α in terms of Z. This is now a reduced one-way classification model, with design matrix $L'A_1$.

Hence the normal equations are

(3.5.28) $$A_1' LL' A_1 \hat{\phi}_\alpha = A_1' LZ$$

$$= A_1' LL' Y^* \quad \text{(by 3.5.19).}$$

But

(3.5.29) $$A_1' LL' A_1 = A_1' A_1 - A_1' A_2 D^{-1}_{n_{.j}} A_2' A_1 ,$$

and, since $A_1' A_2 = N$ as defined in notation g) of this section, and $A_1' A_1 = D_{n_{i.}}$, as defined in h) of this section *),

(3.5.29) reduces to

(3.5.30) $A_1' LL' A_1 = P$ (as defined in (3.5.2)).

Also, since $Y^* = LL'Y$, and since LL' is idempotent, the righthand side of (3.5.2) will be

(3.5.31) $$A_1 LL'Y^* = A_1' LL' Y = A_1'[I - A_2 D^{-1}_{n_{.j}} A_2']Y.$$

Since $A_1' Y = T_a$ (defined in j) and $A_2' Y = T_b$ (defined in k) ,

---

*) For more detail, see, e.g., [5].

$$(3.5.32) \quad A_1' LL'Y^* = T_a - N D^{-1}_{n_{\cdot j}} T_b = Q ,$$

as defined in (3.5.2).

This somewhat lengthy argument shows that elimination of nuisance (here B) effects from each observation vector and subsequent performance of a one-way classification analysis on the A effect produces the same adjusted normal equations, hence the same estimates and H matrix as the direct analysis of the two-way classification model. Since E is easily shown to be the same, $E^{-1} H$ will be identical, and hence the discriminant function will be the same in either case. It must be stressed, here, that the result is not altogether obvious. One might argue that, if nuisance parameters are to be eliminated, their best estimates should be used ($\hat{\Phi}_\beta$ in our notation). In fact, however, the simple arithmetic mean of all observations in each B group had to be used, i.e., the unadjusted estimate.

The foregoing argument also shows that the discriminant function of A-effects is the same in terms of $\underline{y}$ and $\underline{y}^*$. Hence an optimum future assignment would be as follows:

(1) Obtain, from the calibration sample, the discriminant function (same in terms of $\underline{y}$ and $\underline{y}^*$) by (3.5.2), (3.5.4), (3.5.9), and (3.5.11).

(2) Obtain, from the calibration sample, the unadjusted (simple) means of all observations in the group B_i $(i = 1,2,3,4)$ on test j. Put these means into the (4×10), $[(s \times p)]$ matrix F_b.

(3) Let the discriminant function vector obtained from (1)be called $\underline{a}$ (with p(10) elements). The three rows of the solution matrix $(\hat{\phi}_\alpha)$ in (3.5.2) are each multiplied by $\underline{a}$, i.e., $\hat{z}_1 = \underline{\hat{\alpha}}_1' \, \underline{a}$; $\hat{z}_2 = \underline{\hat{\alpha}}_2' \, \underline{a}$; $\hat{z}_3 = \underline{\hat{\alpha}}_3' \, \underline{a}$.

These are the adjusted mean composite scores of the three A-groups.

(4) Let the discriminant function vector obtained from (1) be called $\underline{a}$ (with p(10) elements). Let a new candidate have test scores $\underline{y}'$ and belong to group B_i $(i = 1, 2, 3, 4)$. Subract the i'th row vector of F_b from $\underline{y}'$ to obtain the appropriate $\underline{y}^{*\prime}$. Obtain the candidate's composite score $z = \underline{a}' \, \underline{y}^*$.

(5) Assign the candidate to that group A_j whose composite mean score $\hat{z}_j$ (from (3)) is closest to the candidate's composite score z (from (4)).

6. Example - Multiple Regression

This section will deal with a case where concomitant and random variables are present. For simplicity let us assume that each concomitant variable has arithmetic mean zero, and each random variable has expected overall mean equal zero. Then the univariate mulitple regression model will simply be

$$E(y_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_q x_{qi} \; .$$

In matrix notation,

(3.6.1) $E(\underline{y}) = X \, \underline{\beta}$,

where X represents the matrix of scores on the q concomitant variables. As outlined in Chapter II, the corresponding multivariate model is simply obtained by replacing $\underline{y}$ by Y and the parameter vector $\underline{\beta}$ by the parameter matrix $\underset{\sim}{B}$. We will, for simplicity, make standard least-squares assumptions, which, in the multivariate case, generalize to $\mathrm{var}(\underline{y}_i') = \Sigma$, $\mathrm{cov}(\underline{y}_i', \underline{y}_j') = 0$. Then

(3,6,2) $E(Y) = X \, \underset{\sim}{B}$.

We consider the hypothesis of absence of regression, i.e.

(3.6.3) $\underset{\sim}{B} = 0$.

Hence C, according to the general linear hypothesis, is simply the identity matrix, thus

(3.6.4)
$$\begin{aligned} H &= \hat{\underset{\sim}{B}}' C' [C(X'X)^{-1} C']^{-1} C \, \hat{\underset{\sim}{B}} \\ &= \hat{\underset{\sim}{B}}' X' X \, \hat{\underset{\sim}{B}} \; . \end{aligned}$$

The normal equations are

(3.6.5) $X'X \, \hat{\underset{\sim}{B}} = X'Y$,

i.e.,

(3.6.6) $\hat{\underset{\sim}{B}} = (X'X)^{-1} X'Y$.

Thus

$$(3.6.7) \qquad H = Y'X(X'X)^{-1}X'Y .$$

Also

$$(3.6.8) \qquad E = Y'Y = Y'X \hat{\underset{\sim}{B}}$$
$$= Y'Y - Y'X(X'X)^{-1}X'Y .$$

To obtain an interpretation of the general discriminant function in this case we will consider a formally similar case in which the x's are random variables. Then, except for a constant multiplier (n in this case), X'X represents the sample variance-covariance matrix of the x-set, X'Y is the sample covariance matrix between the x-set and y-set, and Y'Y represents the sample variance-covariance matrix of the y-set. Let them be denoted by S_{xx}, S_{xy}, and S_{yy}, respectively. Then except for a constant multiplier,

$$(3.6.9) \qquad E^{-1}H = [S_{yy} - S'_{xy}\, S^{-1}_{xx}\, S_{xy}]^{-1} S'_{xy}\, S^{-1}_{xx}\, S_{xy} .$$

Introducing $S_{yy}\, S^{-1}_{yy}$ after the square bracket and incorporating S_{yy} into the expression in square brackets one obtains

$$(3.6.10) \qquad E^{-1}H = [I - S^{-1}_{yy}\, S'_{xy}\, S^{-1}_{xx}\, S_{xy}]^{-1}\, S^{-1}_{yy}\, S'_{xy}\, S^{-1}_{xx}\, S_{xy}$$
$$= [I - Q]^{-1} Q \text{ , say.}$$

Now, the matrix Q will be recognized as the matrix whose characteristic roots are the squares of canonical correlations (see, e.g., [14]). The eigenvector associated with the largest canonical correlation, i.e., (3.6.11) $Q \underline{b} = \rho^2 \underline{b}$

is used in constructing Hotelling's "most predictable criterion" [11]. The linear combination $\underline{b}'\underline{y}$ is that linear combination of response variables which can be predicted with greatest precision from the x-set. We will show that $\underline{b}$ equals the discriminant function. From (3.6.11) it follows that

(3.6.12) $(I - Q)\underline{b} = (1 - \rho^2)\underline{b}$,

$$(I - Q)^{-1}\underline{b} = \frac{1}{1 - \rho^2}\,\underline{b} ,$$

and, by (3.6.11), $\underline{b} = 1/\rho^2\, Q\underline{b}$, hence

(3.6.13) $(I - Q)^{-1}\, Q\, \underline{b} = [\rho^2/(1 - \rho^2)]\underline{b}$.

Comparison with (3.6.10) shows that the most predictable criterion function $\underline{b}$ is identical to the discriminant function given by $E^{-1}\, H\, \underline{a} = \lambda\, \underline{a}$. The root, λ, equals $\rho^2/(1 - \rho^2)$, where ρ^2 is the square of the (largest) canonical correlation.

7. Summary

These examples indicate the usefulness and interpret the meaning of the general discriminant function in a variety of situations. It reduces to the classical discriminant function between two groups in the simplest case, produces maximum F-statistics in analysis of variance, permits discrimination on marginal effects in cross-classifications, eliminating the other effects, and reduces to the "most predictable criterion" in the multiple regression example. It will be used for the construction of artificial variables, in the attempt to order and select variables, in Chapters IV and V.

IV. REPRESENTATIVE ORDERING OF VARIABLES

1. Statement of the Problem

In experiments involving several response variables it is frequently desirable to find out which variables are the major contributors to an established significant difference between groups. A somewhat controversial point may arise regarding the decision just when a difference has been established as significant. Subjective preferences for cut-off probabilities, such as .05 or .01, make the meaning of "significance" somewhat semantic. On the other hand, if observed differences in a sample could have arisen, by mere chance, in 25% or 50% of all cases, on the assumption of no difference in the population, an interpretation of such sample differences may have as little scientific value as astrological reasoning. If chance suffices for the interpretation of some observed phenomenon, it is futile to look for a more complicated explanation. The proper procedure, in case of doubt, is not to make more numerical manipulations, but to conduct additional experiments to decide the issue. We will thus assume that by some acceptable criterion the null hypothesis of no group differences has been rejected for the whole set of variables, and look for further interpretations in this case only (see examples in Appendix A).

As described in [2, 20] and illustrated in Appendix A, there are several different approaches to testing the null hypothesis in multivariate analysis. We may test the general linear hypothesis, depending on our choice, by either the likelihood-ratio test, using log $|E|$ - log $|H + E|$ as a test statistic, or by another member of the class of union-intersection statistics [20], which employs the largest characteristic root of $E^{-1}H$ as a test statistic, or still another member which employs $tr(E^{-1}H)$ as a statistic [15] (generalization of Hotelling's T_o^2) or, finally, by a technique equivalent to the analysis of covariance, which is called the "step-down procedure" [19, 21] and which employs the ratio of "leading coefficients" in the left-hand side of the forward Doolittle solution of E to that of (H + E). The term "leading coefficient" denotes the first term of the second to last row in each cycle of the Doolittle solution. This very simple numerical method is implicit in the discussions in [19], but needs to be explicitly stated. Under H_o, these ratios have a Beta distribution, and they are all independent of each other (see proof in [21]). This "step-down" method suffers from the disadvantage that it depends on the order in which the variables are listed and, unless there is an underlying logical "order of importance" of the variables, it cannot be recommended for overall testing purposes. It is, however, useful for construction of confidence regions, and for the tracing of significant contributions due to each variable.

From a logical point of view, Roy's characteristic-root test seems to have the greatest appeal for, if we accept the overall multivariate hypothesis, we know that we cannot find any linear combination of the variables which would lead to rejection of the corresponding univariate hypothesis based upon that linear combination. Acceptance then implies that chance alone could have produced the observed differences with appreciably large probability.

In most cases, however, a hypothesis is formulated in such a way that one wishes to reject it, if possible, and most frequently the tests do lead to rejection. There are many methods in univariate analysis which are employed to explain and interpret the observed differences, and each and all of them can be extended to multivariate analysis in the most elementary manner. However, in multivariate analysis we frequently attempt to find an additional answer, viz., how much of the significant departure from the null hypothesis is due to each variable. It is for this purpose that discriminant functions and the step-down method seem to be valuable tools. The determination of order will be made with the aid of the general discriminant function discussed and illustrated in Chapter III. The procedure is best illustrated by an actual example which will be described in the next section.

The principle of representative ordering of variables, in accordance with assumptions 1) and 4) in Section I.2, will be discussed in Section 4 of this chapter. If the discriminant function is based on matrices E and H obtained from the actual observations, the entire response variable will be considered in establishing the (artificial) best discriminator. If, in accordance with the factor analysis model, one assumes a score on a response variable to consist of a "common" part (x, say) and a "unique" part (u), a discriminant function in terms of the "common" part only would be desirable. It will be shown that, while matrices E and H in terms of x (common part) are indeterminate, one can establish uniquely the correlations of the discriminant function in terms of x and each of the observed variables.

The ordering of variables, for the purpose of tracing significant contributions of variables to an established group difference, will be based upon the correlations between each observed variable and the discriminant function. The higher the (absolute value of the) correlation, the more closely will the particular observed variable be related to the artificial best discriminator. Thus, a response variable having a high correlation with the discriminant function will contribute more strongly to the group discrimination than one having a lower correlation. The ordering of variables should thus be by descending absolute values of correlations versus the discriminant function.

If such ordering is based upon the E and H matrices of the observed variables, the individual variables will determine the contribution to significance. If, however, such ordering is based, implicitly, upon the E and H matrices of the "common" variables, discrimination will be in terms of these common factors only. This ordering will be called representative ordering of variables.

2. Description of a Demonstration Study

An explicit illustration has been prepared to clarify the concepts and methods described in the sequel. Thirteen sets of random normal numbers, 60 each, were obtained from tables given in [7], each with mean 3 and variance 1. These thirteen variables were denoted as follows:

Set 1: x_1
Set 2: x_2
Set 3: x_3
Sets 4 to 13: u_1 to u_{10}.

These sets were subdivided into 3 "groups" of 20 "experimental units" each, i.e., each group consists of 20 observation vectors, for all 13 variables.

In group 1, each value of the set x_1 was augmented by 1, and each value of the set x_2 by 3; the remaining sets or "variables" were left unchanged.

In group 2, x_1 was augmented by 2 and x_2 by 2.

In group 3, x_1 was augmented by 3 and x_2 by 1.

Consequently, the three groups were designed in such a way that differences are due to x_1 and x_2 only, and the augmented numbers were chosen so that $x_1 - x_2$ would be the discriminant function, if only x_1 and x_2 were observed.

From the 13 sets or variables, with x_1 and x_2 augmented as described, the following linear combinations were calculated for each of the 60 vectors:

$$z_1 = 2x_1 + u_1$$
$$z_2 = x_1 + u_2$$
$$z_3 = x_1 + 2u_3$$
$$z_4 = 2x_2 + u_4$$
$$z_5 = x_2 + u_5$$
$$z_6 = x_2 + 2u_6$$
$$z_7 = 2x_3 + u_7$$
$$z_8 = x_3 + u_8$$
$$z_9 = x_3 + 2u_9$$
$$z_{10} = x_1 - x_2 + u_{10}.$$

Thus it is seen that the variables z_1 to z_{10} would represent an almost ideal factor analysis "simple structure." Each variable consists of a part which it shares with at least one other variable, and a part which is uncorrelated with the corresponding part of any other variable. The contribution of the common element is large in some variables (z_1, z_4, z_7) and small in others (z_3, z_6, z_9). All group differences are in terms of the common parts x_1 and x_2 only, there is no group difference in terms of the common part x_3 and none in terms of the specific parts u_1 and u_{10}.

In the sequel, $z_1, z_2, \ldots, z_{10}$ will be treated as the only "observable" variates, and the structure which underlies these variates is assumed to be unknown. The problem of analyzing and finding this structure can be solved by the customary factor analysis techniques. Our problem is to find a method which will trace the significance of group differences to variables involving x_1 and x_2 and, if possible, to identify these components.

3. Ordinary Discriminatory Analysis

3.1 Test of Hypothesis and Computation of Discriminator

Table 4.3.1 represents the matrix of sums-of-squares and sums-of-products "within groups", (W), i.e., the diagonal term for the q'th variable is

$$\sum_{i=1}^{N} x_{qi}^2 - 20[\bar{x}_{q1}^2 + \bar{x}_{q2}^2 + \bar{x}_{q3}^2]$$

where the first sum is over all observations on variable q, and $\bar{x}_{q1}$ represents the mean of the observations on variable q in group 1. Analogously, the off-diagonal term in row q and column r is

$$\sum_{i=1}^{N} x_{qi}x_{ri} - 20[\bar{x}_{q1}\bar{x}_{r1} + \bar{x}_{q2}\bar{x}_{r2} + \bar{x}_{q3}\bar{x}_{r3}].$$

Table 4.3.1.

"Within-Groups" Matrix W (= E)

	1	2	3	4	5	6	7	8	9	10
1	281.3613	90.4307	153.5669	15.7010	-6.1254	-31.5804	-14.2835	3.0611	-23.7179	138.2685
2	90.4307	90.9986	63.3779	-2.8068	-13.0649	-21.3890	-9.6290	-9.7455	-3.3677	44.8778
3	153.5669	63.3779	264.6788	-40.2588	-28.5775	-13.8459	17.6938	-15.2572	-21.7513	98.3283
4	15.7010	-2.8068	-40.2588	195.9019	102.5442	61.9117	-37.6534	-35.1742	40.4363	-60.6417
5	-6.1254	-13.0649	-28.5775	102.5442	108.6544	28.2828	4.3324	-1.4200	35.6306	-38.2377
6	-31.5804	-21.3890	-13.8459	61.9117	28.2828	163.9298	-38.5376	-23.8892	55.5501	-50.1753
7	-14.2835	-9.6290	17.6938	-37.6534	4.3324	-38.5376	264.8693	108.9436	112.6205	11.5665
8	3.0611	-9.7455	-15.2572	-35.1742	-1.4200	-23.8892	108.9436	118.0043	76.1805	12.0840
9	-23.7179	-3.3677	-21.7513	40.4363	35.6306	55.5501	112.6205	76.1805	286.9824	-38.1295
10	138.2685	44.8778	98.3283	-60.6417	-38.2377	-50.1753	11.5665	12.0840	-38.1295	121.1899

Table 4.3.2 is the table of means, for each variable, and for each of the three groups.

In Table 4.3.3, the overall mean for each variable was subtracted from each group mean. This was done in order to simplify calculation of the discriminant function.

Table 4.3.4 represents the inverse of W, each term multiplied by 100.

Instead of performing the tedious multiplication $W^{-1}B$ and obtaining the vector associated with its largest root one may proceed as follows: If the (10 x 3) matrix reproduced in Table 4.3.3 is denoted by G, the matrix of sums-of-squares and -products "between groups," B, is (since $n_1 = n_2 = n_3 = 20$)

$$B = 20GG'. \tag{4.3.1}$$

Thus

$$W^{-1}B = 20\ W^{-1}GG'$$

and

$$\mathrm{ch}(W^{-1}B) = 20\ \mathrm{ch}(G'W^{-1}G), \tag{4.3.2}$$

where ch(A) stands for "characteristic roots of A." The last equality, (4.3.2), refers to the non-zero roots only. Notice that the matrix in (4.3.2) is of order (3 x 3) and of rank 2. The matrix $100\ G'W^{-1}G$ is given in Table 4.3.5.

Table 4.3.2.

Means

Variable $(\underline{x})$	Group I $\bar{\underline{x}}_1$	Group II $\bar{\underline{x}}_2$	Group III $\bar{\underline{x}}_3$
1	9.65170	13.35155	15.03485
2	6.45705	7.91880	9.26730
3	9.47740	10.64345	12.32860
4	15.23150	13.48150	10.56110
5	9.08130	7.96160	6.56410
6	12.00570	11.13145	9.27800
7	9.31340	7.75725	8.65835
8	5.56220	5.61230	6.34300
9	9.01645	8.96105	9.77460
10	0.17010	2.43400	5.25525

Table 4.3.3.

Adjusted Means

Variable $(\underline{x})$	Group I $\bar{\underline{x}}_1 - \bar{\bar{\underline{x}}}$	Group II $\bar{\underline{x}}_2 - \bar{\bar{\underline{x}}}$	Group III $\bar{\underline{x}}_3 - \bar{\bar{\underline{x}}}$
1	-3.02767	.67218	2.35548
2	-1.42400	.03775	1.38625
3	-1.33908	-.17303	1.51212
4	2.14013	.39013	-2.53027
5	1.21230	.09260	-1.30490
6	1.20065	.32640	-1.52705
7	.73707	-.81908	.08202
8	-.27697	-.22687	.50383
9	-.23425	-.28965	.52390
10	-2.44968	-.18578	2.63547

Table 4.3.4.

Inverse of "Within" Matrix (x 100)

$100 \quad W^{-1}$

	1	2	3	4	5
1	1.657973	-.589764	-.300009	-.823233	.079884
2	-.589764	1.777669	-.121752	.020665	.195564
3	-.300009	-.121752	.681927	.206527	-.054103
4	-.823233	.020665	.206527	1.704298	-1.125780
5	.079884	.195564	-.054103	-1.125780	1.980213
6	-.054742	.226296	-.124064	-.148117	.074513
7	.196679	.026502	-.166060	-.027212	-.073521
8	-.386790	.259554	.293136	.483413	-.213965
9	.008137	-.160818	-.004944	-.091906	-.011557
10	-1.816849	.200101	-.145941	1.125784	-.002629

	6	7	8	9	10
1	-.054742	.196679	-.386790	.008137	-1.816849
2	.226296	.026502	.259554	-.160818	.200101
3	-.124064	-.166060	.293136	-.004944	-.145941
4	-.148117	-.027212	.483413	-.091906	1.125784
5	.074513	-.073521	-.213965	-.011557	-.002629
6	.879443	.155958	.106112	-.219627	.298254
7	.155958	.723086	-.544560	-.171087	-.140261
8	.106112	-.544560	1.773652	-.312290	.102526
9	-.219627	-.171087	-.312290	.572577	.141330
10	.298254	-.140261	.102526	.141330	3.675958

Table 4.3.5.

$100G'W^{-1}G$

8.400749	-.278246	-8.122509
-.278246	1.055663	-.777441
-8.122509	-.777441	8.899980

The characteristic equation of this matrix is

$$\lambda^3 - 18.356{,}392\lambda^2 + 26.373{,}249{,}763\lambda = 0$$

and the roots are

$$\lambda_1 = 16.785{,}169 \qquad \lambda_2 = 1.571{,}223 \qquad \lambda_3 = 0.$$

Hence the largest root of $W^{-1}B$ equals $16.785{,}169/5 = 3.35703$. Then

$$(4.3.3) \qquad \theta = \frac{c}{1 + c} = \frac{3.35703}{4.35703} = .7705$$

has the standard distribution of the largest characteristic root with parameters

$$s = 2$$
$$m = (|3 - 10 - 1| - 1)/2 = 3.5,$$

and

$$n = (60 - 3 - 10 - 1)/2 = 23.$$

Heck's tables [10], Chart I list $\theta = .41$ as significant at the .01 level for the stated parameters. Hence the observed θ is very clearly significant and indicates a very significant difference between the three groups. It is, therefore, justifiable to construct a discriminant function.

The discriminant function is the vector associated with the largest root of $W^{-1}B$. Let $\underline{u}$ be the vector (of arbitrary length) associated with the largest root of $G'W^{-1}G$. Then

$$G'W^{-1}G\underline{u} = \lambda\underline{u} \ , \tag{4.3.4}$$

$$W^{-1}GG'(W^{-1}G\underline{u}) = \lambda(W^{-1}G\underline{u}) \ ,$$

$$W^{-1}B(W^{-1}G\underline{u}) = 20\lambda(W^{-1}G\underline{u}) \ , \tag{4.3.5}$$

(since $20GG' = B$), and hence $W^{-1}G\underline{u} = \underline{a}$ is a vector associated with the largest root of $W^{-1}B$, i.e., the discriminant function.

Using $\lambda = \lambda_1$ based upon $100G'W^{-1}G$ in Table 4.3.5 (a scalar multiplier does not change the discriminant function) we found

$$\underline{u}' = [.967{,}690; \ \ .032{,}308; \ \ -1.000{,}000]$$

where the last component was arbitrarily set equal to -1. Hence $\underline{a}' = k\ \underline{u}'G'W^{-1}$ represents a discriminant function which, if applied to the ten z-variables, maximizes the F-ratio between the three groups. Choosing $k = -100$, we obtain

$$\begin{aligned} \underline{a}' = [&.480{,}973; \ \ 1.340{,}384; \ \ -.887{,}045; \\ &-2.399{,}763; \ \ .744{,}483; \ \ -.571{,}121; \ \ -1.166{,}406; \\ &-.492{,}134; \ \ 1.633{,}771; \ \ 3.282{,}312\ ] \ . \end{aligned} \tag{4.3.6}$$

3.2 Interpretation of Descriminant Function

We may regard the discriminant function as an artificial variable which is a linear combination of the observable variables (z). In order to describe the properties of such an artificial variable, it is helpful to have estimates of its correlations versus the observable ones. Denoting this artificial variable by x, we have

(4.3.7) $x = \underline{a}'\underline{z}$.

Since $\mathrm{var}(\underline{z}) = \Sigma$,

(4.3.8) $\mathrm{var}(x) = \underline{a}'M\underline{\alpha}$

and

(4.3.9) $\mathrm{cov}(x,\underline{z}) = \underline{a}'\Sigma$,

where $\mathrm{cov}(x,\underline{z})$ denotes the row-vector of covariances of the artificial variable x and each of the observable variables z_i.

Hence

(4.3.10) $\mathrm{corr}(z_i, x) = (\underline{a}'\Sigma)_i \,/\, \sqrt{\underline{a}'\Sigma\underline{a}}\,\sqrt{\sigma_{ii}}$

is the correlation of the i'th observable versus the artificial variable (where $(\underline{a}'\Sigma)_i$ denotes the i'th component of the row-vector $\underline{a}'\Sigma$ and σ_{ii} is the i'th diagonal term of Σ). We may, in (4.3.10) multiply numerator and denominator by an arbitrary constant, k, and thus, effectively, replace Σ by $k\Sigma$ without changing the correlation. We thus obtain a maximum-likelihood estimate of the correlations by the formula

$$(4.3.11) \qquad r(z_i,\, x) = (\underline{a}'W)_i \,/\, \sqrt{\underline{a}'W\underline{a}}\ \sqrt{w_{ii}}$$

where W represents the "within-groups" matrix reproduced in Table 4.3.1. Using $\underline{a}'$ in (4.3.6) and W in Table 4.3.1 we obtain

$$(4.3.12) \qquad r' = [.55027;\ .50791;\ 30326;\ -.57805;\ -.41737;\ -.36971;\ -.06516;\ .12578;\ .07866;\ .79836]$$

This vector of correlations indicates that the artificial variable, x, which discriminates best between the three groups, is closely related to variable z_{10}, somewhat less to z_1, z_2, and z_3, and, with reverse direction, to z_4, z_5, z_6, while it is practically unrelated with z_7, z_8, and z_9. A comparison with the actually underlying structure given in Section IV.2 shows that it is quite satisfactorily represented by this artificial variable. In terms of their contribution to the differences between the groups, the ten response variables can now be ordered according to decreasing values of the absolute quantities in $\underline{r}'$ above, i.e., in the order

10 - 4 - 1 - 2 - 5 - 6 - 3 - 8 - 9 - 7

4. Factor Analysis Discrimination

4.1 Theoretical Background

In factor analysis, we assume that the vector of observable variates $\underline{z}$ (p x 1) can be represented in terms of a set of artificial variates $\underline{x}$ (k x 1), (where k is (usually much) smaller than p - 1) each of which is shared by at least two observable ones, and a set of uncorrelated ("unique") variates $\underline{u}$ (p x 1). Thus

(4.4.1) $\underline{z} = A\underline{x} + \underline{u}$

where $\mathrm{var}(\underline{u}) =$ diagonal.

Conversely, we may say [4], that the correlations among the observable variates $\underline{z}$ can be explained in terms of the smaller set $\underline{x}$ in the sense that all partial correlations among the z_i's given $\underline{x}$, are zero. With this model which, unlike that of "principal component analysis," is expressed merely in terms of correlations (var(u) = diagonal is the only essential criterion for this model) we may proceed to find a discriminant function applicable to the common artificial x-variates only. Without loss of generality, we may assume the set $\underline{x}$ to be uncorrelated so that

(4.4.2) $\mathrm{var}(\underline{z}) = \Sigma = AD_{\xi i}A' + D_{\eta i}$

where $D_{\xi i}$ and $D_{\eta i}$ are two diagonal matrices. Hence the matrix of correlations among the $\underline{z}$ variables has the structure

(4.4.3) $\mathrm{corr}(\underline{z}) = \underset{\sim}{R} = D_{1/\sqrt{\sigma_{ii}}}AD_{\xi i}A'D_{1/\sqrt{\sigma_{ii}}} + D_{\gamma i}$,

where $D_{1/\sqrt{\sigma_{ii}}}$ is a diagonal matrix whose elements are the reciprocals of the standard deviations of the z_i's. Letting

(4.4.4) $D_{1/\sqrt{\sigma_{ii}}} A D_{\sqrt{\xi i}} = \Phi$,

we have

(4.4.5) $\underset{\sim}{R} = \Phi \Phi' + D_{\gamma i}$.

The arbitrariness of Φ should be noted for, post-multiplication of Φ by any orthogonal (k x k) matrix will result in a new Φ^* which satisfies (4.4.5) equally well. A maximum-likelihood estimate of Φ can be obtained. Any matrix F (p x k) which satisfies the relation (see [4]),

(4.4.6) $(R - FF')D_{1/1-h^2}F = F$

is a maximum-likelihood estimate of a Φ satisfying (4.4.5). In this formula, $D_{1/1-h^2}$ is a diagonal matrix with elements $1/(1-h_i^2)$ where h_i^2 is the sum of squares of the i'th row in F. The matrix F represents a matrix of sample correlations of the artificial variables $(x_1, x_2, \ldots, x_k)$ versus the observable variables $(z_1, z_2, \ldots, z_p)$, and h_i^2 is the square of the multiple correlation of the i'th observable variate, z_i, versus all artificial ones. It is the only term which remains invariant under change of F.

In each of the k groups,

(4.4.7) $E(\underline{z}_j) = E[A\underline{x}_j] + E(\underline{u}_j)$; $(j = 1, 2, \ldots, k)$.

We now make the assumption that group differences are in terms of the "common" artificial variables only, so that (4.4.7) reduces to

(4.4.8) $E(\underline{z}_j) = E[A\underline{x}_j] + E(\underline{u})$.

Since we may add a constant to each element of the row vector $A\underline{x}_j$ (the same for all groups) and thus, effectively, obtain a new artificial variable x_i^* , we can make, without loss of generality, $E(\underline{u}) = \underline{0}$. The resulting artificial variable x_i^* will then enter (4.4.1) in place of x_i, and correlations are unaffected. Thus

(4.4.9) $E(\underline{z}_j) = AE(\underline{x}_j^*)$

represents the relationship between $\underline{z}$ and $\underline{x}^*$ on the assumption that all group differences are in terms of the common artificial variates only. To obtain an expression for the "between groups" and "within groups" matrices of the artificial variable $\underline{x}^*$, we note that

(4.4.10) $W_z = AW_xA'$

and

$$B_z = AB_xA' \quad ,$$

where W_x is the "within groups" matrix in terms of the artificial variables x_i^*, and W_z is the actually observable "within groups" matrix in terms of the z_i. B_x and B_z are corresponding matrices "between groups". (This is generally true, even for more complicated designs, for note that the hypothesis and error matrices H and E for a set of linear transforms $A\underline{z}$ of the original variates are simply AHA' and AEA' (see, e.g., [20]).

Premultiplying and postmultiplying both equations in (4.4.10) by $D_{1/\sqrt{w_{ii}}}$, a diagonal matrix whose elements are the reciprocals of the square-roots of diagonal elements in the "within-groups" matrix (Table 4.3.1), we find

$$(4.4.11) \qquad D_{1/\sqrt{w_{ii}}} W_z D_{1/\sqrt{w_{ii}}}$$

$$= D_{1/\sqrt{w_{ii}}} A D_{\sqrt{x_i}} D_{1/\sqrt{x_i}} W_x D_{1/\sqrt{x_i}} D_{\sqrt{x_i}} A' D_{1/\sqrt{w_{ii}}}$$

and

$$(4.4.12) \qquad D_{1/\sqrt{w_{ii}}} B_z D_{1/\sqrt{w_{ii}}}$$

$$= D_{1/\sqrt{w_{ii}}} A D_{\sqrt{x_i}} D_{1/\sqrt{x_i}} B_x D_{1/\sqrt{x_i}} D_{\sqrt{x_i}} A' D_{1/\sqrt{w_{ii}}} \; .$$

The left-hand side of (4.4.11) is the sample correlation matrix of the observed (z) variables. The left-hand side of (4.4.12) will be denoted by Q, a matrix which is easily obtained from the observed data (see next section). $D_{\sqrt{x_i}}$ is an unknown diagonal matrix. Since A is a matrix of parameters, unknown but satisfying (4.4.5), we will replace it by its maximum-likelihood estimate (see [4] for more explicit derivations). The maximum-likelihood estimate of σ_{ii} is $(1/N)w_{ii}$, so that

$$D_{1/\sqrt{w_{ii}}} A \hat{D}_{\sqrt{x_i}} = (1/\sqrt{N})F \quad ,$$

where F is a matrix satisfying (4.4.6), and the maximum-likelihood estimate of a Φ satisfying (4.4.5).

Thus, from the right-hand side of (4.4.11),-- the left-hand side is R--,

$$(4.4.13) \qquad R= (1/N)FD_{1/\sqrt{x_i}}W_xD_{1/\sqrt{x_i}}F'$$

and, from (4.4.12),

$$(4.4.14) \qquad Q = (1/N)FD_{1/\sqrt{x_i}}B_xD_{1/\sqrt{x_i}}F' \quad .$$

Premultiplying (4.4.13) and (4.4.14) by F' and post-multiplying by F, inverting (4.4.13) and multiplying we find

$$(4.4.15) \qquad (F'RF)^{-1}(F'QF) = (F'F)^{-1}D_{\sqrt{x_i}}W_x^{-1}B_xD_{1/\sqrt{x_i}}(F'F)$$

provided that $(F'RF)^{-1}$ and $(F'F)^{-1}$ exist. Since both are obtained from sample matrices, and are of order k ($< p-1$), they will exist almost everywhere (for conditions on real-valuedness of elements in F, see [4]).

We thus find

$$(4.4.16) \qquad D_{\sqrt{x_i}}W_x^{-1}B_xD_{1/\sqrt{x_i}} = (F'F)(F'RF)^{-1}(F'QF)(F'F)^{-1} \quad .$$

Because of the indeterminancy of the x_i we cannot, from this expression obtain the eigenvector associated with the largest characteristic root of $W_x^{-1}B_x$ (the desired discriminant function) but we can, of course, obtain the root itself, for

$$(4.4.17) \qquad ch(D_{\sqrt{x_i}}W_x^{-1}B_xD_{1/\sqrt{x_i}}) = ch(W_x^{-1}B_x)$$

$$= ch[(F'RF)^{-1}(F'QF)] \ ,$$

where Q is obtained from the left-hand side of (4.4.12). *)

The following argument will show that, while we cannot obtain the desired eigenvector, $\underline{a}_x$, itself, we can obtain the sample correlationsof the artificial variable $\underline{a}_x'\underline{x}$ versus each of the observable variates z_i and thus, while the "factor score" of each sample unit is, in general, not available, we can interpret the discriminant function among the artificial variates in terms of its correlations with each of the observed ones.

Since F is the matrix of (sample) correlations of $\underline{z}$ versus $\underline{x}$, the covariance matrix will be

$$\text{sample cov}(z,x) = D_{\sqrt{s_{ii}}} F D_{\sqrt{x_i}}$$

where the s_{ii} are the maximum-likelihood estimates of the variances σ_{ii} of the observed variates, and x_i are those of the (unknown) artificial variates (sample values corresponding to parameters denoted by ξ_i in (4.4.3)). Thus

(4.4.18) $$\text{sample cov}(\underline{z}, \underline{a}_x'\underline{x}) = D_{\sqrt{s_{ii}}} F D_{\sqrt{x_i}} \underline{a}_x$$

where $\underline{a}_x$ is the eigenvector associated with the largest root of $W_x^{-1}B_x$, thus

---

*) ch(A) denotes "characteristic roots of A"; ch(AB) = ch(BA), except, perhaps, for zero roots.

(4.4.19) $W_x^{-1}B_x\underline{a}_x = \lambda\underline{a}_x$

$$D_{\sqrt{x_i}}W_x^{-1}B_xD_{1/\sqrt{x_i}}D_{\sqrt{x_i}}a_x = \lambda D_{\sqrt{x_i}}a_x \quad .$$

Denoting $D_{\sqrt{x_i}}\underline{a}_x$ by $\underline{t}$, we find

(4.4.20) $D_{\sqrt{x_i}}W_x^{-1}B_xD_{1/\sqrt{x_i}}\underline{t} = \lambda\underline{t}$,

and hence $\underline{t} = D_{\sqrt{x_i}}\underline{a}_x$ is the eigenvector of the matrix (4.4.16). Taking (4.4.18) and considering that the sample variances of z_i are s_{ii}, we obtain the sample correlations:

(4.4.21a) $r = (1/\sqrt{t't})F\underline{t}$, or, if we specify eigenvectors $\underline{t}$ to have unit length,

(4.4.21b) $\underline{r} = F\underline{t}$.

In these formulas, $\underline{r}$ is the desired maximum-likelihood estimate of the correlations of the artificial variate $\underline{a}_x'\underline{x}$ versus the observable variates z_i.

To interpret the "factor discriminant function", we therefore need to find an F satisfying (4.4.6) and obtain the eigenvector, $\underline{t}$, associated with the largest root of the right-hand side of (4.4.16),

$$(F'F)(F'RF)^{-1}(F'QF)(F'F)^{-1}.$$

An iterative procedure to obtain F was outlined in [4] and Appendix B and an IBM 650 program is available [Appendix C]. A shortcut method for obtaining the vector $\underline{t}$ is outlined in the next section.

4.2 Numerical Procedure

Table 4.4.1 represents the matrix of correlations obtained from W (Table 4.3.1).

This correlation matrix was subjected to a standard centroid solution program (AM 557) and refactored twice (AM 561) so that a good initial guess for the iterative procedure may be obtained, and a preliminary test on the number of factors may be made. Four factors were extracted in this way, and the results are stated in Table 4.4.2. At this stage, a preliminary decision on the number of factors was reached by the following procedure (which is described in [4] and represents a slight modification of Bartlett's test [6], see 6.6.021.4 in Appendix C).

Test of Partial Independence

(Preliminary)

	$-m \ln \lvert R(\underline{z} \mid \underline{x}) \rvert$ *) $= \chi^2_{45}$	Probabilities
Independ.	236.69	$\ll 10^{-6}$
1 Factor	176.44	
2 Factors	128.76	
3 Factors	47.41	.376
4 Factors	31.41	.936

Since the maximum-likelihood solution is identical with the maximum partial determinant (see Appendix B), the probabilities must be somewhat larger after the final solution has been reached. Two factors are quite obviously insufficient to describe the observed dependence, and the hypothesis of three factors is quite plausible, even at this preliminary stage. It was therefore decided to obtain the maximum-likelihood solution for

---

*) $\lvert R(\underline{z} \mid \underline{x}) \rvert$ is the determinant of the sample matrix of partial correlations between observable variates ($\underline{z}$), given the artificial ones ($\underline{x}$), (see Appendices B and C).

Table 4.4.1

Correlations (R)

	1	2	3	4	5	6	7	8	9	10
1	1.	.565154	.562737	.066877	-.035033	-.147047	-.052322	.016799	-.083467	.748786
2	.565154	1.	.408377	-.021022	-.131391	-.175124	-.062022	-.094045	-.020840	.427347
3	.562737	.408377	1.	-.176800	-.168516	-.066471	.066826	-.086331	-.078922	.549017
4	.066877	-.021022	-.176800	1.	.702858	.345481	-.165298	-.231343	.170539	-.393567
5	-.035033	-.131391	-.168516	.702858	1.	.211919	.025538	-.012540	.201777	-.333223
6	-.147047	-.175124	-.066471	.345481	.211919	1.	-.184944	-.171761	.256111	-.355981
7	-.052322	-.062022	.066826	-.165298	.025538	-.184944	1.	.616221	.408483	.064558
8	.016799	-.094045	-.086331	-.231343	-.012540	-.171761	.616221	1.	.413969	.101048
9	-.083467	-.020840	-.078922	.170539	.201777	.256111	.408483	.413969	1.	-.204456
10	.748786	.427347	.549017	-.393567	-.333223	-.355981	.064558	.101048	-.204456	1.

Table 4.4.2

F After Twice-Refactored Centroid

	I	II	III	IV
1	.62298	.64357	.33573	.04601
2	.41293	.36567	.11060	-.10387
3	.53017	.38113	.14615	-.23648
4	-.62265	.46345	.39130	.21764
5	-.51089	.22645	.43662	.36601
6	-.49829	.14272	.16896	-.35083
7	.24461	-.52318	.50206	.05073
8	.26393	-.59228	.49481	.18209
9	-.15038	-.26482	.62999	-.23281
10	.84982	.29289	.03753	.06154

Table 4.4.3

Three-Factor Solution After 5 Iterations

	I	II	III	h^2
1	.63600	.65052	.35346	.95260
2	.40597	.40886	.10045	.34207
3	.52704	.32724	.04380	.38678
4	-.67809	.50656	.47721	.94414
5	-.53147	.21293	.47873	.55698
6	-.39643	.11549	.05424	.17344
7	.20160	-.52354	.48451	.54948
8	.29648	-.57429	.55234	.72279
9	-.12023	-.29242	.50741	.35743
10	.84127	.29769	.05450	.79933

Table 4.4.4.

F After 80 Iterations

Maximum-Likelihood Solution

(Accurate to five places).

	I	II	III	h^2
1	.64260	.65479	.35726	.96932
2	.40403	.41156	.09440	.34154
3	.52206	.33428	.02768	.38506
4	-.68823	.51279	.48532	.97214
5	-.52848	.21053	.47196	.54637
6	-.38689	.11534	.05075	.16556
7	.19250	-.51102	.46660	.51591
8	.30107	-.58691	.57226	.76259
9	-.11625	-.28831	.49706	.34371
10	.83229	.29716	.05408	.78394

Table 4.4.5.

Standardized Means (M)

Variable	$(\bar{x}_{1i}-\bar{\bar{x}}_i)/\sqrt{w_{ii}}$	$(\bar{x}_{2i}-\bar{\bar{x}}_i)/\sqrt{w_{ii}}$	$(\bar{x}_{3i}-\bar{\bar{x}}_i)/\sqrt{w_{ii}}$
1	-.180500	.040073	.140426
2	-.149277	.003957	.145320
3	-.082309	-.010636	.092945
4	.152905	.027873	-.180779
5	.116302	.008884	-.125185
6	.093775	.025493	-.119268
7	.045289	-.050328	.005040
8	-.025497	-.020885	.046380
9	-.013828	-.017098	.030926
10	-.222524	-.106876	.239400

three factors. After 5 iterations by formula (4.4.6) the matrix reproduced in Table 4.4.3 was obtained.

Comparison with Table 4.4.2 shows the slow rate of change, which is also observed in Appendix B. It took eighty (!) iterations to produce a solution satisfying (4.4.6) to 5 places. The solution after 80 iterations is reproduced in Table 4.4.4 and represents the maximum-likelihood solution. The final test for the number of factors, based upon this F, is reproduced below.

Test of Partial Independence

(Final)

	$-m \ln\lvert R(\dot{\underline{z}}\mid\dot{\underline{x}})\rvert = \chi^2_{45}$	Probabilities
2 Factors	176.10	$\ll 10^{-6}$
3 Factors	19.03	.99973

The extremely sharp turning point of this "indicating sequence" is typical of every study undertaken thus far ([4] and Appendix B), and leaves absolutely no doubt regarding the appropriate number of factors. The conclusion here is: The probability that two artificial variates could have produced a sample in which the observed dependence ocurred is so small as to make this assumption untenable. The probability that three artificial variates (factors) suffice for this purpose is so large that there can be no doubt regarding the adequacy of this assumption.

Next we proceeded to obtain the components of the matrix expression on the right-hand side of (4.4.16) based upon the maximum-likelihood F.

For the purpose of obtaining F'QF, where Q is the left-hand side of (4.4.12), it is helpful to prepare a matrix M which is computed by dividing each row of G (Table 4.3.3) by the square root of w_{ii} (diagonal entries in Table 4.3.1). This matrix is reproduced in Table 4.4.5. Thus F'QF = 20(F'M)(M'F) requires a quite simple and small-size matrix operation.

The matrix 1/20 F'QF is reproduced below in Table 4.4.6

Table 4.4.6.

1/20 F'QF

.784559	.149102	-.042989
.149102	.040221	-.009991
-.042989	-.009991	.002635

The matrix F'RF was obtained by an IBM 650 program and is reproduced below in Table 4.4.7.

Table 4.4.7.

F'RF

7.543942	1.249272	-.314149
1.249272	3.811197	-.404075
-.314149	-.404075	2.434475

Now, instead of obtaining the matrix expression in (4.4.16) directly we made use of the following procedure which greatly reduced computational work:

Recall that we wish to have the eigenvector associated with the largest characteristic root of

$$(F'F)(F'RF)^{-1}(F'QF)(F'F)^{-1}.$$

Calling this vector $\underline{t}$ we notice that

(4.4.22) $(F'RF)^{-1}(F'QF)(F'F)^{-1}\underline{t} = \lambda(F'F)^{-1}\underline{t}$,

and, with $\underline{t}^* = (F'F)^{-1}\underline{t}$, we find

(4.4.23) $(F'RF)^{-1}(F'QF)\underline{t}^* = \lambda\underline{t}^*$.

To obtain $\underline{t}^*$ (which would require obtaining the largest root of a non-symmetric matrix) we make another simplification: In the process of inverting F'RF by the Doolittle method we may obtain, as a simple by-product, the triangular (lower-triangular) matrices $\tilde{T}$ and $\tilde{T}^{-1}{}'$, which satisfy the relations

(4.4.24a) $F'RF = \tilde{T}\tilde{T}'$,

and

(4.4.24b) $(F'RF)^{-1} = (\tilde{T}')^{-1}\tilde{T}^{-1}$.

These matrices are reproduced in Tables 4.4.8 and 4.4.9 below.

Table 4.4.8.

$\tilde{T}$

2.746624	0	0
.454839	1.898504	0
-.114376	-.185437	1.544994

Table 4.4.9.

$\tilde{T}^{-1}$

.364083	0	0
-.087226	.526730	0
.016484	.063220	.647252

Now, (4.4.23) can be written as

$$(\tilde{T}')^{-1}\tilde{T}^{-1}(F'QF)\underline{t}^* = \lambda t^* \tag{4.4.25}$$

$$\tilde{T}^{-1}(F'QF)(\tilde{T}')^{-1}\tilde{T}'\underline{t}^* = \lambda\tilde{T}'\underline{t}^* \quad ,$$

or

$$\tilde{T}^{-1}(F'QF)\ (\tilde{T}')^{-1}\underline{u} = \lambda\underline{u} \quad , \tag{4.4.26}$$

where $\underline{u} = \tilde{T}'\underline{t}^*$.

Thus, we merely have to obtain the largest root and eigenvector in the symmetric (3 x 3) matrix (4.4.26). This matrix is reproduced in Table 4.4.10 below (the scalar multiplier is irrelevant and originates from Table 4.4.6).

Table 4.4.10.

$1/20\ \tilde{T}^{-1}(F'QF)(\tilde{T}')^{-1}$

.103998	.003678	-.001990
.003678	.003428	-.000296
-.001990	-.000296	.000054

The roots of this matrix are

$$\lambda_1 = .104,171 \quad \lambda_2 = .003,308 \quad \lambda_3 = 0,$$

and $\underline{u}'$, with the last element (arbitrarily) equated to 1 is

$$\underline{u}' = [-\ 52.048916;\ \ -1.903242;\ \ 1.000000].$$

Then, from (4.4.26)

$$(\underline{t}^*)' = \underline{u}'T^{-1} = [-18,76763;\ -.93927;\ .64725]$$

and, from (4.4.22)

(4.4.27) $\underline{t} = (F'F)\underline{t}^*$.

The matrix F'F is reproduced below in Table 4.4.11,

Table 4.4.11.

F'F

2.585272	.258530	-.071564
.258530	1.807486	-.065425
-.071564	-.065425	1.393370

and (4.4.27) becomes

$$\underline{t}' = [-48.80858;\ -6.59206;\ 2.30640]\ .$$

Formula (4.4.21b) requires that $\underline{t}$ be reduced to unit length; thus

$$\underline{t}' = [.98992;\ \ .13370;\ \ -.04678].$$

Then we obtain the final result, i.e., the correlation between the artificial variable which discriminates best between the three groups in terms of common factors only and each of the observed variates, z_1 to z_{10}, as

$$\underline{t}'F' = \underline{r}' = [.70696;\ .45057;\ .56020;\ -.63544;\ -.51708;\ -.36994;\ .10041;\ .19280;\ -.17688;\ .86110].$$

As in the ordinary discriminant function this vector describes adequately the underlying structure. Again it shows that the artificial "discriminator variable" is correlated most highly against z_{10}; then z_1, z_2, z_3; and (with negative sign) z_4, z_5, z_6; and practically uncorrelated with z_7, z_8, z_9. The order of descending absolute values,

10 - 1 - 4 - 3 - 5 - 2 - 6 - 8 - 9 - 7 ,

is, in this case, essentially equal to that of the ordinary discriminatory analysis. The correlations of the artificial variable (discriminant variable) and the observed variables (z) is somewhat higher in the factor analysis approach.

The detailed description of the numerical procedure has been made in order to show how a seemingly complicated and tedious computational procedure may be, by some elementary algebra, reduced to several very simple and small-sized operations.

5. Step-Down Analysis and Interpretaion of Results

The leading coefficients of the forward Doolittle solution (first term in second to last row - before normalizing to unity each cycle) for the within matrix W, and for the matrix B + W (the matrix of sums-of-squares and products corrected for grand mean only) are given below in Table 4.5.1 (for ordering according to results in factor analysis discrimination). The ratio of any two corresponding terms is a Beta variable, and the associated probability is stated below the value of the ratio.

Table 4.5.1.

Step-Down Analysis

Var. No.	10	1	4	3	5	2	6	8	9	7
W	121.1899	123.6076	107.2598	167.5092	54.2190	59.6188	128.5596	100.2191	187.9361	138.2962
B + W	380.8129	147.0071	121.5128	177.6182	54.4400	63.4845	129.0946	101.8586	206.6229	147.0414
Beta	.31824	.84083	.88270	.94309	.99594	.93911	.99586	.98390	.90956	.94053
Prob.	$<10^{-6}$*	.0078*	.0324*	.2055	.8978	.1953	.8995	.6665	.0980	.2296

The starred entries indicate significant contributions. The analysis showsthat, after explaining the contribution due to variable z_{10}, there is a significant contribution left due to z_1 and z_4. In other words, the analysis shows, correctly, that the discrimination can best be explained in terms of variable 10 (which, actually, corresponds to the ideal discriminant function $x_1 - x_2$) and in terms of variables 1 and 4 (each is, in fact, that variable which contains the relatively largest amount of the common factors x_1 and x_2, respectively. It shows that, given the variables z_{10}, z_1, and z_4, there is no further significant contribution due to any additional variable, in complete agreement with the underlying structure. Thus, the "representative" factor analysis discriminant functions indicate an ordering which is in agreement with the (here known) actually underlying order. The same result, however, would have been found from the ordinary discriminant analysis. It should be noted that an arbitrary ordering would have produced significant results for another set of variables, -- for example, if variables z_3 and z_6 had been placed first and second, they would, of necessity, contribute significantly, before the really strong representatives, like z_1 and z_4 are introduced. It is, therefore, important to have some objective criterion for the ordering of variables; in the present case we chose the idea of an underlying structure as such a criterion and obtained very satisfactory results.

Since z_{10} is so obviously the variable with the largest contribution to the discriminant function (it had been, as will be recalled, designed to be the discriminant function itself) the step-down analysis was repeated once again, this time omitting variable z_{10}. The results are stated in Table 4.5.2.

This table again shows, even more clearly than the other one, that the variates which (as we know in the present case) contain the strongest relative contribution to the common factors x_1 and x_2 are, under this ordering, the only ones explaining the observed significant differences between groups.

6. Conclusions and Recommendations

A possibility of statistical interpretation of significant group differences in multivariate analysis is available by the use of discriminant functions and subsequent step-down analysis. If the differences, instead of being due to the observed variables, are due to latent artificial variables which are shared by two or more of the observable ones, a factor analysis discrimination analysis will reveal this fact. It may be somewhat surprising that the ordinary discriminant

Table 4.5.2.

Omitting z_{10}

Var. No.	1	4	3	5	2
W	281.3613	195.0257	168.6372	54.2420	59.8549
B + W	584.6992	329.8070	179.3639	54.5778	63.5320
Beta	.48121	.59133	.94020	.99385	.94212
Prob.	$<10^{-6}$*	$<10^{-6}$*	.1834	.8465	.2060

	6	8	9	7
W	134.7415	100.4270	189.0750	139.3274
B + W	137.3168	101.8952	206.7046	147.9060
Beta	.98125	.98559	.91471	.94200
Prob.	.6113	.6907	.1077	.2313

analysis correctly revealed the hidden structure, even though many of the observed variables were contaminated with very large "unique" contributions which would tend to camoflage the structure among the latent "common" variables. It should be recalled, though, that in the present demonstration study there was actually no mean difference between the unique variables.

According to the theory, a mean difference in the unique variables should produce an entirely different structure in the ordinary discriminant analysis, and should reveal essentially the same structure as before in the factor analysis discriminatory analysis.

Practical usefulness of the factor analysis discriminatory method would be expected in those cases (notably in anthropology and psychology) where many observations are made on available and measurable traits, but where differences between, e.g., ethnic groups or educational systems, are postulated in terms of some underlying classificatory variables only, of which the available measurements reflect a more or less substantial part.

The factor analysis approach is indicated by assumptions 1 and 4 in Section I.2., and the foregoing example illustrates this principle of representative ordering.

V. REPRESENTATIVE SELECTION OF VARIABLES

1. Statement of the Problem

The purpose of this chapter is a study of the role of the discriminant function as a basis for selecting variables. Let us assume that we have k groups of experimental units, and on each experimental unit we make p observations. We would like to find that linear combination of the variables which produces a new variable such that the F-ratio between the k groups is a maximum. This linear combination of the variables is the (estimate of the) general discriminant function. We may regard this variable as an artificial variable constructed from the observed ones. We may then proceed to obtain the correlations of this artificial variable versus each of the observed variables. This vector of correlations will give us an idea of the proximity of the discriminant function to each of the observable variables. If, for example, we find a few correlations extremely high, say about .80, we can say that there is hardly any difference between the best discriminator and those particular observed variables. We can then take the variables which have high correlations versus the discriminant function as almost equivalent in discriminatory power to that of the discriminant function itself. If, in further experimentation, we intend to reduce the number of variables on which we would like to perform measurements, we will of course select those variables which have the highest correlations (in absolute value) versus the best discriminator.

The best way to introduce the concept of *representative* selection is in the form of an example. Suppose we are interested in the difference between human beings of different races. We are not only interested in the physical characteristics but also, let us say, in their physiological differences, such as blood pressure, certain chemicals in the blood stream, respiration rate, etc. Let us assume, then, that we have three different racial groups and that we make 10 anthropological measurements, such as general height, weight, arm length, color of skin, etc., and that we also make 10 measurements on physiological characteristics such as blood pressure, content of calcium in the blood stream, etc. If we apply the previously mentioned discriminant function as a principle of selection, we will only include physical or anthropological variables in our selected set, because, naturally, the racial groups differ most strikingly in terms of physical characteristics. Thus, if we should perform a new set of experiments based only on the reduced set of variables, we would measure nothing but physical characteristics. This would completely obliterate any physiological measurements that we have previously conducted. It is in this context that the concept of *representative* selection enters. We may wish to select subsets out of the whole set of variables such that all the important common characteristics of the original set of variables are retained. It is thus necessary to classify the whole set of variables into certain subsets that belong together.

The classical method of finding subsets of variables such that they are classified under certain consistent principles of classification, is factor analysis. In factor analysis, or, more generally, in Dependence Analysis, an attempt is made to explain the observed dependence among all the variables in terms of a few outside or artificial variables. These variates, incidentally, are not linear combinations of the original variables. (This is a very striking difference between factor analysis and what is known as component analysis, which has no bearing on the present problem.) The method of classification of subsets of variables is explained in detail in Section V.2, where the various preferred solutions of factor analysis are explained and defined, and where the role of the "Simple Structure" as a classification principle is outlined. It is then shown that representative selection is performed in two stages. In the first stage, any group differences which may exist are eliminated in that we consider only the matrix of sums-of-squares and products due to error. From this matrix we construct a correlation matrix and on the basis of this correlation matrix we classify the variables into subsets.

In Section V.2, the role of the discriminant function as a basis for the selection of variables, and the principle of representative selection by factor analysis is explained in detail. Section V.3 presents a detailed demonstration study, where a factorial structure of the variables is known. It consists of a sampling experiment in which the variables have been arranged so as to form a structure in three artificial variables or factors. Section V.4 presents a study of actual data which are an excerpt from a study on mentally retarded children, carried out in 1957-60 (see [24]). We selected three different sets of variables from these; intelligence measures, straight achievement measures, and measures of achievement gain in the course of one year. We were interested in discriminating between two groups: those children who stayed in the public school system, and those students who were transferred to special classes for mentally retarded children. Previous analysis of these data showed that there was a rather striking interaction between the difference in school types and the age of the subjects, in that older children in the public education system performed considerably better than the older children that had been referred to special education classes. The difference was not nearly as great in the younger age groups. For that reason, we had to make three separate studies in investigating the discriminant functions and in finding the best representatives among the observed variables. A study was made for the children 11 years and younger, between 11 and 13 years of age, and for the group 13 years of age and older. The rather interesting and somewhat surprising findings of the studies are mentioned in Section V.4.

In the actual analysis of these data one must make use of algebraic advantages by representing matrices in the smallest possible form. The details of numerical evaluations in this case are presented in Section V.3.

For clarification it must be pointed out that the method of selecting variables on the basis of their correlations on the best discriminator does not insure that one will find subsets of 2, 3, or more variables which produce maximum F-ratios of all possible doublets, triplets, etc. No method for the selection of the best discriminating single variable can be used for finding, in general, the best two or three, for the "best" doublet may not even contain the best single variable. For illustration, let us consider the following example:

Let z_1 and z_2 be two uncorrelated variables, and let $z_3 = z_1 + z_2 + u$, where u is independent of z_1 and z_2 and has the same variance. Suppose we found, with z_1 and z_2 expressed in standard units, that $d = z_1 + z_2$ is the best discriminator between the groups under study in the sense that this combination of the three variables (z_1, z_2, z_3)

would produce the highest F-ratio between groups. Thus, the correlations would be

	z_1	z_2	z_3	d
z_1	1.000	0.000	.577	.707
z_2	0.000	1.000	.577	.707
z_3	.577	.577	1.000	.816

Clearly, z_3 is the best single discriminating variable (it is, except for a relatively small error u, equal to the discriminant function itself, even though on the "partial" regression weights of d on z_3 given z_1 and z_2 it need not appear at all). However, the best pair of variables is z_1 and z_2 and does not even contain the best single variable z_3. In fact, the multiple correlations of d on every possible pair are

$$R\ (d;\ z_1,\ z_2) = 1.000$$

$$R\ (d;\ z_1,\ z_3) = .866$$

$$R\ (d;\ z_2,\ z_3) = .866$$

so that z_3 is never a member of the best pair.

In the method proposed in this chapter the discriminant function is regarded as some unknown latent variable which produces maximum discrimination, and observable variables are selected in terms of their closeness to this ideal underlying variable. In this sense, it is natural to ask whether just one ideal variable would be sufficient to explain the discrimination, and it is for this reason that the representative selection has been proposed as an extension of this idea.

2. Theory of Representative Selection

As in Chapter IV we will assume that each experimental unit has p observations, that by some design and hypothesis matrix the experimental units fall into groups, and that, from a calibration sample (see Chapter III), a matrix E has been obtained and some matrix H (there may be different possibilities for formulating H in a given calibration sample, but there is only one E; see Appendix A). As indicated in Chapter III, a discriminant function, $\underline{a}$, will be the eigenvector associated with the largest root of $E^{-1} H$. If the response variables are denoted by $\underline{z}$, an artificial variable, $u = \underline{a}'\underline{z}$, will be called the discriminant function or "best discriminator". This artificial variable u can be correlated with each of the observed variables, $z_1, z_2, \cdots z_p$. As shown in (4.3.11) the sample correlation between u and $z_1, z_2, \cdots z_p$ is the row vector

$$(5.2.1) \qquad \underline{r}' = (1/\sqrt{\underline{a}'E\,\underline{a}})\underline{a}'E\ D_{1/\sqrt{e_{ii}}}\ ,$$

where $D_{1/\sqrt{e_{ii}}}$ is a diagonal matrix whose principal elements are the reciprocals of the square-roots of the diagonal elements in E. The elements of this vector are, as in (4.3.11) computed as

$$(5.2.2) \qquad r_i = (\underline{a}'E)_i / \sqrt{\underline{a}'E\,\underline{a}}\,\sqrt{e_{ii}} \quad ,$$

where $(\underline{a}'E)_i$ denotes the i'th element in the row vector $\underline{a}'E$.

Given a set of p observable variables, $z_1, z_2, \ldots, z_p$, we want to subdivide them into subsets. In other words, we want to find some principle of classification which enables us to put certain variables into a common class. This task is usually performed by Factor Analysis (see [4], [12], and [23]). The purpose of Factor Analysis is the study of dependence patterns in the variables. This is accomplished by seeking artificial variables which may explain the dependence among the observable variables. It is well known, in the theory of partial correlations, that dependence in a set of variables may be due to the presence of a common outside variable. The goal in Factor Analysis is to find one or more artificial variables which could explain the dependence of all the observable variates.*) These artificial variables are described in terms of their correlations with the observable ones and these correlations have been called "factor loadings". The artificial variables, $\underline{y}$, must be determined in such a way that they "explain the dependence" in the set $\underline{z}$.

---

*) There is the further attempt, analogous to the problem of fitting polynomials (see [4]), to keep the number of such artificial variables or factors as small as possible.

The indeterminacy of F, the matrix of sample correlations of the observable versus the artificial variables, was discussed in Chapter IV. Unlike the problem in Chapter IV, however, where all we needed was some solution to equation (4.4.6), we must in this case consider some preferred solution. Various types of preferred solutions are discussed in Appendix B, and they may all have some significance depending upon the problem.

As pointed out in [4] this problem, in all its aspects, is not unlike the problem of fitting polynomials. The coefficients, γ_i, of the equation $y_i = \gamma_0 + \gamma_1 P_1(x_i) + \ldots + \gamma_k P_k(x_i)$, where $P_j(x_i)$ denotes the j'th orthogonal (Tchebychev) polynomial of x, may be easy to obtain, mathematically, but they are certainly difficult to interpret, physically. For example, let y_i be the position of a mass falling in a vacuum. Here, the γ's in $y_i = \gamma_0 + \gamma_1 P_1(t_i) + \gamma_2 P_2(t_i)$ are without meaning; but if we rewrite the right-hand side (without changing the function) in the representation $y_i = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2$, α_0 can be interpreted as the position at time 0, α_1 can be interpreted as the initial speed, and α_2 as $\frac{1}{2}g$, where g is the acceleration due to gravity. This may serve as an illustration of the fact that different representations of the same function may serve different purposes.

The same is true of the representation of artificial variables in factor analysis. A transformation of the arbitrary matrix F will always result in a matrix whose elements are correlations of the artificial variables (combined and represented in some linear form) versus the observed ones.

For simplicity of interpretation, the Simple Structure concept has been proposed by Thurstone [23]. The Simple Structure is obtained by finding individual transformation vectors such that, when pre-multiplied by F, they yield other vectors with as many zero loadings (correlations) as possible and just a few high loadings. If such a vector or vectors can be found, one can assign the variables to subsets. For we then have an artificial variable closely associated with a few of the observable ones and unrelated with the rest. Hence, the observable variables with the high loadings must contain some "factor" which does not influence the others. It is clear, then, that the variables with high loadings in such a vector should be assigned to a common subset or class. Ideally, when we have k factors, we would like to find just k such vectors of the simple structure. This would insure that all k artificial variables, which have been shown to be required to explain the dependence, are susceptible to interpretation. This ideal, however, is not always reached.

Once a classification of variables into subsets has been found, the problem of representative selection of variables for the purpose of discrimination can be solved by applying the general discrimination techniques to each subset of variables. Numerical shortcuts, especially in the evaluation of eigenvectors, will be presented in each case.

3. An Artificial Demonstration Study

3.1 Description of Data

This explicit demonstration has been constructed in order to investigate the possibilities of detecting underlying order in a sample. Fifteen sets of random normal numbers, 90 each, were obtained from Table A-23 of [7]. Table A-23 is a table of random normal numbers with mean 2 and variance 1, but each number taken from the table was augmented by 3 to produce random normal numbers with mean 5 and variance 1. From one given line in the table the first fifteen numbers were extracted to make one observation vector, and the first ninety lines were chosen to constitute ninety observation vectors.

These fifteen variables were denoted as:

Set 1: x_1
Set 2: x_2
Set 3: x_3
Sets 4-15: $u_1, u_2, \ldots u_{12}$.

These sets were further subdivided into 3 groups of 30 "experimental units" each, i.e., each group consisted of 30 observation vectors. Group effects were introduced into the variable x_1, x_2, x_3, u_3, u_5, and u_8, in the following manner:

In group 1, each value of the set x_1 was augmented by 6, x_2 by 0, x_3 by 2, u_3 by 4, u_5 by 2, and u_8 by 6; the remaining sets were unchanged.*

In group 2, each value of the set x_1 was augmented by 3, x_2 by 2, x_3 by 0, u_3 by 2, u_5 by 1, and u_8 by 6; the remaining sets were unchanged.

In group 3, each value of the set x_1 was augmented by 0, x_2 by 4, x_3 by 1, u_3 by 0, u_5 by 0, and u_8 by 0; the remaining sets were unchanged.

Now using these augmented variables, a new set of variables was constructed as follows:

$$
\begin{aligned}
z_1 &= 3x_1 + u_1 \\
z_2 &= 5x_1 + u_2 \\
z_3 &= x_1 + u_3 \\
z_4 &= 2x_1 + u_4 \\
z_5 &= 6x_2 + u_5 \\
z_6 &= 3x_2 + u_6 \\
z_7 &= 2x_2 + u_7 \\
z_8 &= x_2 + u_8 \\
z_9 &= x_3 + u_9 \\
z_{10} &= 4x_3 + u_{10} \\
z_{11} &= x_1 + x_2 + u_{11} \\
z_{12} &= x_2 + 6x_3 + u_{12}
\end{aligned}
$$

---

*) Variables augmented by zero are, of course, unchanged. The difference in designation was made to distinguish the six variables in which a change was made in some group.

$z_1, z_2, \ldots, z_{12}$ will be used as the "observable" variables for the study. It will be noted that this leads to a demonstration of a three-factor study, with x_1, x_2, and x_3 as common factors and the uncorrelated u's as specifics. (The population parameters of the z variables are presented in Table 5.3.1.) Thus, the analysis will be made on the z variables only, and the structure of these variables will be assumed unknown. This will enable us to compare the results of the analysis with the actual underlying structure later on. The technique of finding the unknown structure will be the usual Factor Analysis method.

3.2 Preparation of Data

The matrix of sums-of-squares and sums-of-products within groups, E, is given in Table 5.3.2 and E^{-1} is given in Table 5.3.3. The constructed z variables, the matrix E, and its inverse E^{-1} were obtained on the I.B.M. 650. The elements of the matrix of estimated correlations R were obtained as

$$r_{ij} = \frac{e_{ij}}{\sqrt{e_{ii}}\,\sqrt{e_{jj}}} \quad .$$

The matrix R is given in Table 5.3.4, and R^{-1} in Table 5.3.5.

Table 5.3.1

Population Means and Variances*

	Group I	Group II	Group III
z_1	38, 10	29, 10	20, 10
z_2	60, 26	45, 26	30, 26
z_3	20, 2	15, 2	10, 2
z_4	27, 5	21, 5	15, 5
z_5	37, 37	48, 37	59, 37
z_6	20, 10	26, 10	32, 10
z_7	15, 5	19, 5	23, 5
z_8	16, 2	15, 2	14, 2
z_9	12, 2	10, 2	11, 2
z_{10}	33, 17	25, 17	29, 17
z_{11}	21, 3	20, 3	19, 3
z_{12}	52, 38	42, 38	50, 38

*An entry in the table of μ, σ^2 means that the variable is distributed normal with mean μ and variance σ^2.

Table 5.3.2

Matrix E

	1	2	3	4
1	783.2361	1153.8684	244.1520	470.9815
2	1153.8684	1987.4942	407.7200	771.9945
3	244.1520	407.7200	161.2637	149.1448
4	470.9815	771.9945	149.1448	389.9145
5	216.7007	227.3985	6.0731	87.0861
6	106.3894	117.4361	2.6563	43.3284
7	19.2648	23.9384	- 21.3434	23.2976
8	19.8597	20.7101	- 18.6569	13.7184
9	- 31.2640	- 38.0216	- 10.3488	5.7388
10	- 97.9325	- 152.7416	- 16.5333	- 20.2888
11	243.8801	416.7521	71.2785	169.6193
12	- 81.7115	- 113.3096	7.3703	12.5446

	5	6	7	8
1	216.7007	106.3894	19.2648	19.8597
2	227.3985	117.4361	23.9384	20.7101
3	6.0731	2.6563	- 21.3434	- 18.6569
4	87.0861	43.3284	23.2976	13.7184
5	2924.1632	1465.9236	881.6001	406.0168
6	1465.9236	802.0511	440.4736	206.1355
7	881.6001	440.4736	372.4584	136.2462
8	406.0168	206.1355	136.2462	148.1918
9	7.2575	- 4.1125	10.4524	7.6397
10	- 253.9752	- 142.1466	- 62.2690	- 39.6836
11	494.6827	251.1353	153.8364	83.1061
12	173.6893	68.0215	93.9487	17.6623

	9	10	11	12
1	- 31.2640	- 97.9325	243.8801	- 81.7115
2	- 38.0216	- 152.7416	416.7521	-113.3096
3	10.3488	- 16.5333	71.2785	7.3703
4	5.7388	- 20.2888	169.6193	12.5446
5	7.2575	- 253.9752	494.6827	173.6893
6	- 4.1125	- 142.1466	251.1353	68.0215
7	10.4524	- 62.2690	153.8364	93.9487
8	7.6397	- 39.6836	83.1061	17.6623
9	155.3080	250.3869	- 2.8052	386.2087
10	250.3869	1078.1841	- 50.6908	1484.6407
11	- 2.8052	- 50.6908	240.5619	50.9893
12	386.2087	1484.6407	50.9893	2415.9008

Table 5.3.3

Inverse of E (x 100) $100\ E^{-1}$

	1	2	3	4
1	1.048036	- .467815	- .114314	- .338727
2	- .467815	.563363	- .294057	- .322011
3	- .114314	- .294057	1.442965	.137537
4	- .338727	- .322011	.137537	1.293669
5	- .151980	.075709	- .048972	.093598
6	.065273	- .031236	.014043	- .007524
7	.205905	- .031564	.137828	- .196459
8	- .017160	.012133	.216205	- .009795
9	.115241	- .002437	- .064943	- .170075
10	- .070300	.025528	.134557	- .016213
11	.117531	- .290285	.078825	- .114159
12	.039160	.003720	- .121298	.007088

	5	6	7	8
1	- .151980	.065273	.205905	- .017160
2	.075709	- .031236	- .031564	.012133
3	- .048972	.014043	.137828	.216205
4	.093598	- .007524	- .196459	- .009795
5	.547898	- .756221	- .311455	- .078865
6	- .756221	1.507099	.042809	- .043135
7	- .311455	.042809	1.069861	- .102457
8	- .078865	- .043135	- .102457	1.191245
9	- .056918	.080648	.031569	- .088476
10	.050679	.026324	.103608	.002519
11	- .123679	- .031557	- .070042	- .203996
12	- .026777	- .017077	- .080821	.018440

Table 5.3.3 (Cont.)

	9	10	11	12
1	.115241	- .070300	.117531	.039160
2	- .022437	.025528	- .290285	.003720
3	- .064943	.134557	.078825	- .121298
4	- .170075	- .016213	- .114159	.007088
5	- .056918	.050679	- .123679	- .026777
6	.080648	.026324	- .031557	- .017077
7	.031569	.103608	- .070042	- .080821
8	- .088476	.002519	- .203996	.018440
9	1.129534	- .081994	.088817	- .127282
10	- .081994	.880481	.098820	- .540000
11	.088817	.098820	1.301669	- .097665
12	- .127282	- .540000	- .097665	.402893

Table 5.3.4

Matrix of Correlations R

	1	2	3	4	5	6	7	8
1	1.000000	.924819	.686982	.852261	.143190	.134230	.035668	.058293
2	.924819	1.000000	.720179	.876954	.094327	.093014	.027823	.038161
3	.686982	.720179	1.000000	.594779	.008844	.007386	- .087087	- .120687
4	.852261	.876954	.594779	1.000000	.081557	.077479	.061135	.057070
5	.143190	.094327	.008844	.081557	1.000000	.957215	.844757	.616781
6	.134230	.093014	.007386	.077479	.957215	1.000000	.805898	.597992
7	.035668	.027823	- .087087	.061135	.844757	.805898	1.000000	.579927
8	.058293	.038161	- .120687	.057070	.616781	.597992	.579927	1.000000
9	- .089640	- .068435	- .065392	.023321	.010769	- .011652	.043459	.050358
10	- .106570	- .104342	- .039650	- .031291	- .143035	- .152858	- .098262	- .099278
11	.561845	.602714	.361890	.553831	.589811	.571733	.513933	.440157
12	- .059401	- .051612	.011808	.012925	.065348	.048866	.099040	.029519

Table 5.3.4 (Cont.)

	9	10	11	12
1	- .089640	- .106570	.561845	- .059401
2	- .068435	- .104342	.602714	- .051612
3	- .065392	- .039650	.361890	.011808
4	.023321	- .031291	.553831	.012925
5	.010769	- .143035	.589811	.065348
6	- .011652	- .152858	.571733	.048866
7	.043459	- .098262	.513933	.099040
8	.050358	- .099278	.440157	.029519
9	1.000000	.611883	- .014513	.630501
10	.611883	1.000000	- .099533	.919889
11	- .014513	- .099533	1.000000	.066885
12	.630501	.919889	.066885	1.000000

Determinant of R = .0000033889622

Table 5.3.5

Matrix R^{-1}

	1	2	3	4
1	8.208512	- 5.836711	.406284	- 1.871866
2	- 5.836711	11.196745	- 1.664739	- 2.834741
3	- .406284	- 1.664739	2.326971	.344873
4	- 1.871866	- 2.834741	.344873	5.044218
5	- 2.299995	1.825134	- .335983	.999417
6	.517384	- .394414	.050189	- .042063
7	1.112115	- .271547	.337777	- .748696
8	- .058538	.065904	.334230	- .023539
9	.401938	- .124661	.102772	- .418528
10	- .645977	.373719	.561058	- .105179
11	.510165	- 2.007187	.155259	- .349638
12	.538625	.081499	- .757098	.068846

	5	6	7	8
1	- 2.299995	.517384	1.112115	- .058538
2	1.825134	- .394414	- .271547	.065904
3	- .335983	.050189	.337777	.334230
4	.999417	- .042063	- .748696	- .023539
5	16.020433	-11.580732	- 3.250564	- .517588
6	-11.580732	12.087967	.234184	- .150351
7	- 3.250564	.234184	3.984793	- .240745
8	- .517588	- .150351	- .240745	1.765369
9	- .383667	.284755	.075933	- .134266
10	.899803	.244841	.656583	.010051
11	- 1.037676	- .138248	- .209648	- .385151
12	- .711569	- .237859	- .766676	.110353

Table 5.1.5 (Cont.)

	9	10	11	12
1	.401938	- .645977	.510165	.538625
2	- .124661	.373719	-2.007187	.081499
3	.102772	.561058	.155259	- .757098
4	- .418528	- .105179	- .349638	.068846
5	- .383667	.899803	-1.037676	- .711569
6	.284755	.244841	- .138248	- .237859
7	.075933	.656583	- .209648	- .766676
8	- .134266	.010051	- .385151	.110353
9	1.754263	- .335519	.171683	- .779670
10	- .335519	9.493215	.503281	-8.715259
11	.171683	.503281	3.131309	- .744553
12	- .779670	- 8.715259	- .744553	9.733502

3.3 Factor Analysis

The Factor Analysis performed in this study was done in order to determine the structure. The rotation to Simple Structure will enable us to divide the variables into representative sets. Once this structure has been determined, the discriminant function for each set of variables may be calculated and in turn one may find the correlations of the observable variables versus the discriminant function (artificial variable). These correlations will then be the basis for ordering the observable variables.

The Complete Factor Analysis was computed on the I.B.M. 650 by Factor Analysis programs (Appendix C).

Table 5.3.6 represents an improved centroid solution which was used as input for the maximum-likelihood iterations and for the purpose of making a preliminary decision regarding the number of factors. This preliminary test is given in Table 5.3.7. (See also Appendices B and C).

Clearly, two factors are insufficient to explain the observed dependence. The decrease in the X^2-value is rather marked but, according to the preliminary test, three factors are not sufficient. Four factors are clearly sufficient. It must be noted, however, that the X^2-value will be smaller, and hence the associated probability larger, once the maximum-likelihood solution is available. Following the preliminary decision, 4 factors were used for the iteration.

Table 5.3.6

Improved Centroid Solution for F

	1	2	3	4	5
1	.639272	.353586	.617691	-.031142	-.019289
2	.648319	.344169	.651531	.062182	.065902
3	.417864	.200558	.565298	-.177402	.026378
4	.630414	.242303	.581188	.203596	-.109026
5	.668936	.286984	-.644712	-.178239	-.088879
6	.638772	.295459	-.624540	-.192338	-.055062
7	.564108	.200344	-.626736	.016578	-.024671
8	.434837	.176577	-.493017	.199836	.023676
9	.258480	-.628913	-.044823	.151889	-.136842
10	.245689	-.911991	.077838	-.057242	.036091
11	.765070	.316069	-.026707	.096924	.180986
12	.432299	-.875119	-.033021	-.094667	.100788

Table 5.3.7

Tests of Partial Independence (see Appendix B)

(Preliminary)

	χ^2_{66}	Probability
Independence	1034.89	$< 10^{-6}$
1 Factor	977.05	$< 10^{-6}$
2 Factors	698.43	$< 10^{-6}$
3 Factors	114.34	.0001
4 Factors	62.56	.60
5 Factors	49.80	.93

into a maximum-likelihood solution. The procedure was stopped when all communalities agreed to four places. This required 35 iterations. The test for partial independence was repeated yielding, for 4 factors, a χ^2 of 13.105 (66 d.f.) corresponding to a normal curve equivalent of 6.33, which denotes an extremely high probability ($> 1-10^{-9}$). There may thus be a possibility that three factors would be quite sufficient. The maximum-likelihood solution for three factors was also established (after 45 iterations starting from centroid) and, finally, also that for 2 factors (after 10 iterations starting from the principal-axes form of the three-factor solution). The χ^2-values for 3 and 2 factors, respectively, were $\chi^2 = 28.702$ (66 d.f.), normal curve equivalent 3.87, corresponding to a probability of .99995; $\chi^2 = 247.58$ (66 d.f.), normal curve equivalent -10.81, corresponding to a probability less than 10^{-20}. Hence the final indicating sequence (see [4]) is, for the probability of k factors being sufficient,

$$\alpha_2 < 10^{-20}$$

$$\alpha_3 = .99995$$

$$\alpha_4 > 1 - 10^{-9} \quad ,$$

and there can be no doubt about the fact that three factors is the correct number. Thus, the test correctly identifies the known underlying degree of dependence. The (arbitrary) maximum-likelihood solution F for three factors is shown in Table 5.3.8. This was rotated into the Principal-Axes form* which is shown in Table 5.3.9. For interpretation of results, a Simple Structure (see [23]) solution was sought which is shown in Table 5.3.10. If there were no sampling errors the Simple Structure would look as shown in Table 5.3.11. The values of the theoretical structure presented in Table 5.3.11 were obtained by the following argument:

Consider the variable, $z_i = a\,x_1 + b\,x_2 + c\,x_3 + u_i$, then

$$(5.3.1) \qquad h_i^2 \text{ (communality)} = \frac{\text{var}(z_i - u_i)}{\text{var}(z_i)} \quad .$$

Now, the x_1 loading is $a \cdot k$, x_2 loading $b \cdot k$, and x_3 loading $c \cdot k$, where

$$(5.3.2) \qquad (a \cdot k)^2 + (b \cdot k)^2 + (c \cdot k)^2 = h_i^2 \quad ;$$

e.g., consider $z_{12} = x_2 + 6x_3 + u_{12}$:

$$h_{12}{}^2 = 37/38 \quad ,$$

$$k^2 + 36k^2 = 37/38 \ , \quad \text{or} \quad k^2 = 1/38 \quad ,$$

$$k = 1/\sqrt{38} = .162 \quad .$$

Hence the x_1 loading is 0, x_2 loading .162 and x_3 loading .973 (6 x .162).

A comparison of Table 5.3.10 (observed Simple Structure) with Table 5.3.11 (theoretical Simple Structure based upon the population parameters) shows the good agreement of all values, and hence demonstrates the usefulness of the method of analysis, even for a comparatively small sample.

---

*) i.e., an F satisfying (4.4.6) and for which F'F = diagonal.

Table 5.3.8

Maximum-Likelihood Solution

F Matrix after 45 Iterations (3 Factor)

	I	II	III
1	.635837	.353673	.594324
2	.640844	.347614	.665759
3	.450229	.195895	.538746
4	.603770	.255958	.605164
5	.681370	.287370	-.663668
6	.653979	.292243	-.643477
7	.565200	.194770	-.607847
8	.410116	.183341	-.433747
9	.240975	-.596202	-.024891
10	.254358	-.919182	.079711
11	.754218	.301190	.012109
12	.434745	-.883480	-.035051

Table 5.3.9

Principal Axes (3 Factor)

	I	II	III	h_i^2
1	.762427	.545839	-.057958	.882594
2	.768341	.615496	-.074623	.974752
3	.521373	.498714	-.103841	.531328
4	.688834	.552083	-.130340	.796276
5	.694425	-.710350	.021893	.987302
6	.673851	-.686854	.036257	.927157
7	.551983	-.649625	-.012881	.726864
8	.420767	-.460643	.026622	.389946
9	-.070372	-.101467	-.631584	.414145
10	-.204377	-.026255	-.934606	.915947
11	.806310	-.048369	-.085040	.659707
12	-.035835	-.157335	-.971973	.970769

Table 5.3.10

Simple Structure 3 Factors

Var.	A	B	C	h_i^2	Belonging to Subset
1	.927	.062	-.041	.8826	A
2	.981	.011	-.030	.9748	A
3	.729	-.059	.025	.5313	A
4	.890	.005	.036	.7963	A
5	.019	.989	-.008	.9873	B
6	.017	.959	-.023	.9272	B
7	-.037	.849	.031	.7269	B
8	-.011	.624	-.016	.3899	B
9	-.017	-.010	.641	.4141	C
10	-.014	-.174	.945	.9159	C
11	.563	.544	.034	.6597	A and B
12	.026	.031	.982	.9708	C

Table 5.3.11

Expected Simple Structure (3 Factors)

	x_1	x_2	x_3	h_i^2
1	.949	.000	.000	.900
2	.981	.000	.000	.962
3	.707	.000	.000	.500
4	.894	.000	.000	.800
5	.000	.986	.000	.973
6	.000	.949	.000	.900
7	.000	.894	.000	.800
8	.000	.707	.000	.500
9	.000	.000	.707	.500
10	.000	.000	.970	.941
11	.577	.577	.000	.667
12	.000	.162	.973	.974

Incidentally, if we had taken the overfactored study (4 factors) as a starting point, the same result would have shown in the Simple Structure. Table 5.3.12 shows the Simple Structure formed from the four factor solution. A fourth overdetermined plane*) could not be found.

As expected the Simple Structure produced three representative sets and they are as follows:

Set I - Variables 1, 2, 3, 4, and 11
Set II - Variables 5, 6, 7, 8, and 11
Set III - Variables 9, 10, and 12 .

3.4 Discriminatory Analysis

The discriminant function $\underline{a}$ is given by the expression

$$E^{-1} H \underline{a} = \lambda \underline{a} \quad , \tag{5.3.3}$$

and the solution for $\underline{a}$ is the eigenvector associated with the largest characteristic root of $E^{-1} H$, where E^{-1} is the matrix given in Table 5.3.3 and H is the matrix for the hypothesis of no group differences. Hence the elements must be sums-of-squares and products "between groups", or

$$h_{ij} = \sum_{r=1}^{3} n_r (\bar{x}_r^{(i)} - \bar{\bar{x}}^{(i)})(\bar{x}_r^{(j)} - \bar{\bar{x}}^{(j)}) \quad , \tag{5.3.4}$$

where $\bar{x}_r^{(p)}$ is the mean of the p'th variable in the r'th group, and $\bar{\bar{x}}^{(p)}$ is the mean of the p'th variable over all 3 groups.

---

*) Vector with sufficiently large number of small entries.

Table 5.3.12

Simple Structure 4 Factors

Var.	A	B	C	Belonging to Subset
1	.914	.066	-.037	A
2	.970	.007	-.032	A
3	.695	-.040	.036	A
4	.902	-.008	.029	A
5	.004	.992	.000	B
6	.005	.952	-.016	B
7	-.023	.828	.025	B
8	.017	.596	-.026	B
9	-.007	-.017	.637	C
10	-.017	-.172	.947	C
11	.577	.522	.026	A and B
12	.022	.029	.978	C

Now, for ease of calculation let B B' = H, where the (i, r)'th element of B is $(\bar{x}_r^{(i)} - \bar{\bar{x}}^{(i)})\sqrt{n_r}$, and n_r is the size of group r. Thus the matrix B in Set I mentioned above would be:

$$\sqrt{30}\begin{bmatrix} (\bar{x}_1^{(1)} - \bar{\bar{x}}^{(1)}) & (\bar{x}_2^{(1)} - \bar{\bar{x}}^{(1)}) & (\bar{x}_3^{(1)} - \bar{\bar{x}}^{(1)}) \\ (\bar{x}_1^{(2)} - \bar{\bar{x}}^{(2)}) & (\bar{x}_2^{(2)} - \bar{\bar{x}}^{(2)}) & (\bar{x}_3^{(2)} - \bar{\bar{x}}^{(2)}) \\ (\bar{x}_1^{(3)} - \bar{\bar{x}}^{(3)}) & (\bar{x}_2^{(3)} - \bar{\bar{x}}^{(3)}) & (\bar{x}_3^{(3)} - \bar{\bar{x}}^{(3)}) \\ (\bar{x}_1^{(4)} - \bar{\bar{x}}^{(4)}) & (\bar{x}_2^{(4)} - \bar{\bar{x}}^{(4)}) & (\bar{x}_3^{(4)} - \bar{\bar{x}}^{(4)}) \\ (\bar{x}_1^{(11)} - \bar{\bar{x}}^{(11)}) & (\bar{x}_2^{(11)} - \bar{\bar{x}}^{(11)}) & (\bar{x}_3^{(11)} - \bar{\bar{x}}^{(11)}) \end{bmatrix}$$

using variables 1, 2, 3, 4, 11. These calculations were further eased by utilizing the following relations:

$\bar{x}_r^{(i)} = \frac{T_r^{(i)}}{30}$ where $T_r^{(i)}$ is the total of the i'th variable in group r; $\bar{\bar{x}}^{(i)} = \frac{G^{(i)}}{90}$ where $G^{(i)}$ is the total of the i'th variable over all three groups.

Now, by Equation (5.3.3),

(5.3.5) $\quad E^{-1} B B' \underline{a} = \lambda \underline{a}$ and

(5.3.6) $\quad B' E^{-1} B[B' \underline{a}] = \lambda[B' \underline{a}]$.

Let $B' \underline{a} = \underline{u}$, hence

(5.3.7) $\quad B' E^{-1} B \underline{u} = \lambda \underline{u}$.

We find the largest characteristic root of the small symmetric matrix $B' E^{-1} B$ and the eigenvector associated with the root is $\underline{u}$. (Note that $E^{-1} H$ would be non-symmetric, 12 x 12, whereas $B' E^{-1} B$ is symmetric, 3 x 3.) Then from Equation (5.3.7),

(5.3.8) $\quad E^{-1} B B' E^{-1} B \underline{u} = \lambda E^{-1} B \underline{u}$ or

(5.3.9) $\quad E^{-1} H[E^{-1} B \underline{u}] = \lambda[E^{-1} B \underline{u}]$, hence

(5.3.10) $\quad \underline{a} = E^{-1} B \underline{u}$

and the desired discriminant function is easily found.

This procedure was followed in obtaining the discriminant function for each of the three sets derived from the Simple Structure and also for the set of all twelve variables taken together. The task was performed as follows:

(a) The matrix B was computed;

(b) The matrices $E^{-1} B$ and $B' E^{-1} B$ were computed;

(c) Characteristic roots and associated eigenvectors of $B' E^{-1} B$ were obtained;

(d) The largest characteristic root and associated eigenvector $\underline{u}$ was chosen;

(e) Using $\underline{u}$ and $E^{-1} B$ from above, $E^{-1} B \underline{u}$ was computed.

The final step gave $\underline{a}$.

Set I (Variables 1, 2, 3, 4, 11)

The matrix B is given in Table 5.3.13 and E in Table 5.3.14. The discriminant function is

$$\underline{a}' = [.0828, .2670, 1.6459, .3809, -.9859]$$

or in terms of the z_i's ,

$$\underline{a}'\underline{z} = [828_z z_1 + 2,670\ z_2 + 16,459\ z_3 + 3,809\ z_4 - 9,859\ z_{11}].$$

Table 5.3.13

Matrix B Set I (Variables 1, 2, 3, 4, 11)

	1	2	3
1	49.8840	1.2101	-51.0941
2	81.3360	4.1348	-85.4709
3	25.6413	2.4560	-28.0973
4	31.8943	2.7882	-34.6825
11	4.5844	1.4907	- 6.0750

Table 5.3.14

Matrix E Set I (Variables 1, 2, 3, 4, 11)

	1	2	3	4	11
1	783.2361	1153.8684	244.1520	470.9815	243.8801
2	1153.8684	1987.4942	407.7200	771.9945	416.7521
3	244.1520	407.7200	161.2637	149.1448	71.2785
4	470.9815	771.9945	149.1448	389.9145	169.6193
11	243.8801	416.7521	71.2785	169.6193	240.5619

Set II (Variables 5, 6, 7, 8, 11)

The matrix B is given in Table 5.3.15 and E in Table 5.3.16. The discriminant function is

$$a' = [.0946,\ .7880,\ .8568,\ -2.0283,\ -1.1757]$$

or in terms of the z_i's ,

$$\underline{a}'\underline{z} = [946\ z_5 + 7{,}880\ z_6 + 8{,}568\ z_7 - 20{,}283\ z_8 - 11{,}757\ z_{11}] .$$

Set III (Variables 9, 10, 12)

The matrix B is given in Table 5.3.17 and E in Table 5.3.18. The discriminant function is

$$\underline{a}' = [.1671,\ .5778,\ -.3398]$$

or in terms of the z_i's ,

$$\underline{a}'\underline{z} = [1{,}671\ z_9 + 5{,}778\ z_{10} - 3{,}398\ z_{12}] .$$

Set of all variables

The matrix B is given in Table 5.3.19 and E in Table 5.3.1. The discriminant function is

$$\underline{a}' = [-.2547,\ -.1163,\ -2.1608,\ -.1431,\ .1376,\ .6685,\ .1992,\ -2.6915,\ -.1175,\ -.3834,\ .5770,\ .2301]$$

and

$$\underline{a}'\underline{z} = [-\ 2547\ z_1 - 1163\ z_2 - 21{,}608\ z_3 - 1431\ z_4 + 1376\ z_5 + 6685\ z_6 + 1992\ z_7 - 26{,}915\ z_8 - 1175\ z_9 - 3834\ z_{10} + 5770\ z_{11} + 2301\ z_{12}] .$$

Table 5.3.15

Matrix B Set II (Variables 5, 6, 7, 8, 11)

	1	2	3
5	-57.6623	4.8951	52.7672
6	-31.9302	2.4607	29.4696
7	-21.7481	2.2249	19.5232
8	5.5677	.2918	- 5.8595
11	4.5844	1.4907	- 6.0750

Table 5.3.16

Matrix E Set II (Variables 5, 6, 7, 8, 11)

	5	6	7	8	11
5	2924.1632	1465.9236	881.6001	406.0168	494.6827
6	1465.9236	802.0511	440.4736	206.1355	251.1353
7	881.6001	440.4736	372.4585	136.2462	153.8364
8	406.0168	206.1355	136.2462	148.1918	83.1061
11	494.6827	251.1353	153.8364	83.1061	240.5619

Table 5.3.17

Matrix B Set III (Variables 9, 10, 12)

	1	2	3
9	4.1024	- 6.1467	2.0442
10	15.4846	-18.4680	2.9834
12	13.0248	-29.1103	16.0855

Table 5.3.18

Matrix E Set III (Variables 9, 10, 12)

	9	10	12
9	155.3080	250.3869	386.2087
10	250.3869	1078.1841	1484.6407
12	386.2087	1484.6407	2415.9008

Table 5.3.19

Matrix B All Variables

1	49.8840	1.2101	-51.0941
2	81.3360	4.1348	-85.4709
3	25.6413	2.4560	-28.0973
4	31.8943	2.7882	-34.6825
5	-57.6623	4.8951	52.7672
6	-31.9302	2.4607	29.4696
7	-21.7481	2.2249	19.5232
8	5.5677	.2918	- 5.8595
9	4.1024	- 6.1467	2.0443
10	15.4846	-18.4680	2.9834
11	4.5844	1.4907	- 6.0750
12	13.0248	-29.1103	16.0855

3.5 Ordering

As expressed above we may regard the discriminant function as an artificial variable which is a linear combination of the observable variables. One may interpret the relative contribution of each observable variable to this artificial variable by computing the correlations of the artificial variable versus the observable ones. The correlation r_i of the artificial variable versus an observable one was evaluated using the relation (see Equation (5.2.2))

$$(5.3.11) \qquad r_i = (\underline{a}'\, E)_i / \sqrt{\underline{a}' E\, \underline{a}}\ \ \sqrt{e_{ii}}$$

where $(\underline{a}'\, E)_i$ represents the i'th element of the row vector $\underline{a}'\, E$, and e_{ii} is the (i, i) diagonal element of the matrix E presented in Tables 5.3.1, 5.3.14, 5.3.16, or 5.3.18. Now in vector form $\underline{r}' = [r_1\ r_2\ \dots\ r_p]$, hence

$$(5.3.12) \qquad \underline{r}' = \frac{1}{\sqrt{\underline{a}'\, E\, \underline{a}}}\ \underline{a}'\ E\ D_{1/\sqrt{e_{ii}}} \quad .$$

This $\underline{r}'$ was evaluated for each of the three sets and the set of all observable variables taken together. The results are as follows:

Set I (variables 1, 2, 3, 4, 11)

$$\begin{array}{cccccc} & 1 & 2 & 3 & 4 & 11 \\ \underline{r}' = [& .7669, & .7962, & .9022, & .7186, & .1478] \end{array}$$

Set II (variables 5, 6, 7, 8, 11)

$$\begin{array}{cccccc} & 5 & 6 & 7 & 8 & 11 \\ \underline{r}' = [& .4771, & .5065, & .4999, & -.2186, & -.1592] \end{array}$$

Set III (variables 9, 10, 12)

$$\begin{array}{lccc} & 9 & 10 & 12 \\ \underline{r}' = [& .3929, & .6069, & .2566] \end{array}$$

Set of all variables (1, 2, 3, ..., 12)

$$\begin{array}{lcccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \underline{r}' = [& .5888, & .6109, & .6913, & .5507, & -.3327, & -.3533, \\ & 7 & 8 & 9 & 10 & 11 & 12 \\ & -.3483, & .1533, & .0246, & .0595, & .1126, & -.1295] \ . \end{array}$$

The order will then be determined by the descending order of magnitude of absolute values, thus

Set I : order - (3, 2, 1, 4, 11)
Set II : order - (6, 7, 5, 8, 11)
Set III : order - (10, 9, 12)
Set of all variables : order - (3, 2, 1, 4, 6, 7, 5, 8, 12, 11, 10, 9) .

3.6 Selection

Clearly, from the construction of the variables, z_3 was expected to have the highest discriminating power (the standardized difference between two groups there was $(\mu_1^{(3)} - \mu_2^{(3)})/\sigma^{(3)} = (20 - 15)/\sqrt{2} \doteq 3.5$, and this value is larger than that for any other variable. On the other hand if, e.g., only three variables were to be selected for future purposes, the overall discriminant function would tell us to use 3, 2, 1 which, as the factorial structure shows, are purely representative of one factor only (see discussion on anthropological and physiological measurements in Section V.1). Making use of the fact that the twelve variables fall into three sets, our representative choice would be 3, 6, 10. These are, in fact, the best discriminators in each set.

4. Example of Representative Selection of Variables in an Educational Experiment

4.1 Description of Data

The data used in this study were taken from T. G. Thurstone's study given in [24]. The experimental units in this case were mentally retarded children from both public and special schools. These children were not institutionalized, but are in the lower range of the normal population. Many types of test were taken by these children and the findings are presented in [24]. We have chosen 12 of these tests as our observable variables. They are as follows:

z_1 = Binet Mental Age
z_2 = "Primary Mental Abilities" Test Mental Age
z_3 = Achievement Paragraph Meaning
z_4 = Achievement Word Meaning
z_5 = Achievement Spelling
z_6 = Achievement Arithmetic Reasoning
z_7 = Achievement Arithmetic Computation
z_8 = Gain Paragraph Meaning

z_9 = Gain Word Meaning
z_{10} = Gain Spelling
z_{11} = Gain Arithmetic Reasoning
z_{12} = Gain Arithmetic Computation .

Variables z_1 and z_2 are measured in months and variables $z_3, z_4, \ldots, z_{12}$ are measured in grade equivalents, i.e., a response of 2.0 would indicate performance at the second grade level. The children were placed in three age groups:

young	(10 years 11 months and younger)
middle	(11 years to 12 years 11 months)
old	(13 years and older) .

Previous analysis of these data showed a significant interaction between school types and age groups; the older group performed much better in public schools than either the young or middle group. For this reason we made three separate studies to find the discriminating variables between public and special schools, one study for each age group. The number of observations are shown in Table 5.4.1, a total of 480 experimental units.

Table 5.4.1

Number of Observations

	Young	Middle	Old	Total
Public	80	40	72	192
Special	120	60	108	288
Total	200	100	180	480

The matrix of corrected sums-of-squares and products pooled over the six groups is presented in Table 5.4.2 and is denoted as the matrix E. The matrix of correlations R is presented in Table 5.4.3.

4.2 Factor Analysis

The Factor Analysis performed in this study was done in the same way as in Section V.3 and Appendix B. An improved centroid solution for F was obtained and iterated into the maximum-likelihood estimate of F (25 iterations) with five factors. The final decision was five factors since the probability was approximately .92 (normal curve equivalent of 1.41) indicating a good plausibility of five factors. The maximum-likelihood solution for F after 25 iterations is presented in Table 5.4.4.

Once F was obtained and the decision of the number of factors made, the "Principal Axes" were obtained by rotation, and are presented in Table 5.4.5; the communalities are also given there. The "Simple Structure" was then obtained by rotation for the purpose of identifying the variables belonging to common-factor sets. The Simple Structure is presented in Table 5.4.6. The five sets found were: A, which contains intelligence measures (variables 1, 2, 6, 7); B, which contains achievement measures (variables 3, 4, 5, 6, 7); C, which contains gain measures (variables 8, 9, 10, 11); and the other two factors, D and E, of the Simple Structure are

Table 5.4.2

Matrix E

	1	2	3	4
1	59450.00	49485.00	2088.80	2091.40
2	49485.00	92416.00	2944.20	2864.00
3	2088.80	2944.20	434.90	373.88
4	2091.40	2864.00	373.88	417.79
5	2275.10	3088.60	407.08	421.27
6	2952.40	3786.80	313.93	299.42
7	3229.30	4234.10	281.46	273.10
8	138.40	62.10	- 29.49	23.06
9	127.30	49.90	24.50	- 6.60
10	104.40	357.50	31.65	37.25
11	210.20	42.30	32.08	29.69
12	55.00	214.30	20.07	14.40

	5	6	7	8
1	2275.10	2952.40	3229.30	138.40
2	3088.60	3786.80	4234.10	62.10
3	407.08	313.93	281.46	- 29.49
4	421.27	299.42	273.10	23.06
5	665.19	367.64	350.28	40.34
6	367.64	414.96	373.12	24.75
7	350.28	373.12	554.53	16.91
8	40.34	24.75	16.91	141.25
9	42.30	36.40	33.42	41.48
10	22.90	31.07	29.20	23.53
11	47.50	- 5.70	39.91	18.53
12	22.09	21.60	16.51	8.95

Table 5.4.2 (cont.)

	9	10	11	12
1	127.30	104.40	210.20	55.00
2	49.90	357.50	42.30	214.30
3	24.50	31.65	32.08	20.07
4	- 6.60	37.25	29.69	14.40
5	42.30	22.90	47.50	22.09
6	36.40	31.07	- 5.70	21.60
7	33.42	29.20	39.91	16.51
8	41.48	23.53	18.53	8.95
9	100.61	23.23	18.98	14.12
10	23.23	129.98	12.68	12.29
11	18.98	12.68	131.36	20.46
12	14.12	12.29	20.46	126.27

Table 5.4.3

Matrix of Correlations R

	1	2	3	4
1	1.000000	.667612	.410796	.419645
2	.667612	1.000000	.464407	.460915
3	.410796	.464407	1.000000	.877119
4	.419645	.460915	.877119	1.000000
5	.361786	.393927	.756854	.799114
6	.594424	.611499	.738984	.719116
7	.562432	.591459	.573138	.567388
8	.047760	.017188	- .118983	.094926
9	.052051	.016365	.117125	- .032192
10	.037557	.103149	.133119	.159849
11	.075219	.012140	.134217	.126736
12	.020074	.062733	.085645	.062695

	5	6	7	8
1	.361786	.594424	.562432	.047760
2	.393927	.611499	.591459	.017188
3	.756854	.738984	.573138	- .118983
4	.799114	.719116	.567388	.094926
5	1.000000	.699756	.576740	.131604
6	.699756	1.000000	.777827	.102230
7	.576740	.777827	1.000000	.060421
8	.131604	.102230	.060421	1.000000
9	.163511	.178147	.141489	.347956
10	.077880	.133783	.108763	.173656
11	.160690	- .024414	.147873	.136035
12	.076221	.094363	- .062393	.067016

Table 5.4.3 (cont.)

	9	10	11	12
1	.052051	.037557	.075219	.020074
2	.016365	.103149	.012140	.062733
3	.117125	.133119	.134217	.085645
4	- .032192	.159849	.126736	.062695
5	.163511	.077880	.160690	.076221
6	.178147	.133783	- .024414	.094363
7	.141489	.108763	.147873	- .062393
8	.347956	.173656	.136035	.067016
9	1.000000	.203138	.165099	.125275
10	.203138	1.000000	.097040	.095932
11	.165099	.097040	1.000000	.158863
12	.125275	.095932	.158863	1.000000

Determinant of R = .001001523

Table 5.4.4

Matrix F (Maximum-Likelihood Solution, 25 Iterations)

	1	2	3	4	5
1	.628866	.230941	-.321293	.336145	.043374
2	.638647	.278587	-.265802	.326764	.015320
3	.809300	.284088	.431709	-.040941	-.258013
4	.809083	.321143	.420957	-.112223	.214127
5	.764734	.137118	.245682	-.186324	.054366
6	.870004	.319699	-.224076	-.278563	-.113919
7	.735920	.233001	-.229531	-.024823	-.041500
8	.248250	-.488106	-.166390	-.204546	.456236
9	.354375	-.696355	-.115519	-.123671	-.282028
10	.224312	-.179844	.039434	-.043212	.062182
11	.185520	-.265430	.180950	.199239	.042936
12	.120990	-.104362	.004798	-.063868	-.050182

Table 5.4.5

Principal Axes

	I	II	III	IV	V	h_i^2
1	.643	.079	-.460	.027	.188	.6669
2	.668	.129	-.408	.005	.185	.6631
3	.877	.131	.375	-.238	.085	.9903
4	.880	.106	.381	.240	.074	.9934
5	.787	-.081	.263	.073	-.035	.7017
6	.926	-.017	-.145	-.055	-.344	.9999
7	.763	-.010	-.239	-.019	-.102	.6509
8	.100	-.625	-.063	.414	-.036	.5775
9	.159	-.755	-.023	-.350	.012	.7187
10	.169	-.233	.055	.040	.048	.0899
11	.106	-.223	.099	-.017	.329	.1791
12	.092	-.137	.032	-.057	-.025	.0321

Table 5.4.6

Simple Structure

	A	B	C	D	E
1	.693	-.031	.017	.010	.035
2	.661	.026	-.025	-.008	-.027
3	.011	.758	.006	.065	-.330
4	-.004	.803	-.008	-.306	-.025
5	.023	.659	.134	-.069	.038
6	.372	.495	-.021	-.009	-.034
7	.458	.271	.030	.017	.017
8	.006	.003	.553	.021	.741
9	-.002	-.016	.751	.706	.375
10	-.005	.112	.247	.097	.193
11	.003	.035	.328	.181	.159
12	-.018	.068	.134	.112	.063

Transformation Vectors from F

I	II	III	IV	V
.361	.477	.347	.156	.176
.193	.248	-.935	-.653	-.652
-.709	.654	.015	-.124	-.277
.574	-.532	.053	.160	-.129
.023	.011	.052	-.712	.671

well overdetermined pseudo-factors showing a very high loading on a verbal gain variable and a negative one on the corresponding achievement variable. No such inverse relationship was observed in the arithmetic variables. These two factors, D and E, are probably due to the fact that the children in the lowest group on the achievement tests had scores so low in the first administration that they could not but gain in the second.

4.3 Discriminatory Analysis

The discriminant function between school types was sought for each of the three sets mentioned in Section 4.2, on each of the three age groups. Also, the overall discriminant function for all twelve variables on each of the three age groups was determined.

As presented in the previous chapters, the discriminant vector $\underline{a}$ is the eigenvector associated with the largest characteristic root of $E^{-1} H$, or

$$(5.4.1) \qquad E^{-1} H \underline{a} = \lambda \underline{a} \ .$$

Now, in this case,

$$(5.4.2) \qquad H = \sum_{i=1}^{2} n_i (\overline{\underline{y}}_i - \overline{\overline{\underline{y}}})(\overline{\underline{y}}_i' - \overline{\overline{\underline{y}}}')$$

where $\overline{\underline{y}}_1$ is the mean vector for public schools, and $\overline{\underline{y}}_2$ the mean vector for special schools. This reduces to the form

$$(5.4.3) \qquad H = \frac{n_1 n_2}{n_1 + n_2} (\overline{\underline{y}}_1 - \overline{\underline{y}}_2)(\overline{\underline{y}}_1' - \overline{\underline{y}}_2') \ ,$$

and hence

$$E^{-1} H = \frac{n_1 n_2}{n_1 + n_2} E^{-1} (\bar{y}_1 - \bar{y}_2)(\bar{y}_1' - \bar{y}_2') \quad . \tag{5.4.4}$$

We shall denote $(\bar{y}_1 - \bar{y}_2)$ by $\underline{\bar{d}}$ and $(\bar{y}_1' - \bar{y}_2')$ by $\underline{\bar{d}}'$, hence

$$E^{-1} H = \frac{n_1 n_2}{n_1 + n_2} E^{-1} \underline{\bar{d}}\, \underline{\bar{d}}' \quad . \tag{5.4.5}$$

Since ch(A B) = ch(B A), the characteristic root of $E^{-1} H$ is the same as the characteristic root of $\frac{n_1 n_2}{n_1 + n_2} \underline{\bar{d}}' E^{-1} \underline{\bar{d}}$, which is a scaler quantity, hence

$$\lambda = \frac{n_1 n_2}{n_1 + n_2} \underline{\bar{d}}' E^{-1} \underline{\bar{d}} \tag{5.4.6}$$

the only non-zero characteristic root of E^{-1} H. Now by pre-multiplying Equation (5.4.6) by $E^{-1} \underline{\bar{d}}$ we get

$$\frac{n_1 n_2}{n_1 + n_2} E^{-1} \underline{\bar{d}}\, \underline{\bar{d}}' E^{-1} \underline{\bar{d}} = \lambda E^{-1} \underline{\bar{d}} \quad , \tag{5.4.7}$$

or

$$E^{-1} H(E^{-1} \underline{\bar{d}}) = \lambda(E^{-1} \underline{\bar{d}}) \quad . \tag{5.4.8}$$

Hence,

$$\underline{a} = E^{-1} \underline{\bar{d}} \quad . \tag{5.4.9}$$

This vector $\underline{a}$, which can, of course, be multiplied by any arbitrary constant, is equivalent to the discriminant function for two groups as given by R. A. Fisher (see [8]).

The matrix E for the young group is presented in Table 5.4.7, for the middle group in Table 5.4.8, and the old group in Table 5.4.9. The mean vectors for both public

Table 5.4.7

Matrix E for Young (age: up to 10 yrs. 11 mo.)

	1	2	3	4	5	6	7	8
1	16929.00	17437.00	789.50	820.60	877.10	837.10	1006.10	-30.60
2	17437.00	31297.00	1210.90	1257.00	1417.40	1372.70	1600.40	79.90
3	789.50	1210.90	127.30	117.09	136.21	86.43	93.90	-7.27
4	820.60	1257.00	117.09	131.77	140.10	83.60	94.64	5.48
5	877.10	1417.40	136.21	140.10	217.45	114.30	119.21	11.59
6	837.10	1372.70	86.43	83.60	114.30	106.07	99.61	11.63
7	1006.10	1600.40	93.90	94.64	119.21	99.61	148.88	9.92
8	-30.60	79.90	-7.27	5.48	11.59	11.63	9.92	42.73
9	-56.00	-27.50	-.11	-7.50	8.45	8.59	6.46	21.89
10	58.70	125.40	5.92	11.80	-.33	6.33	9.90	10.09
11	138.90	178.60	16.41	17.26	20.70	4.21	24.11	5.48
12	6.70	111.90	2.64	4.54	8.14	5.69	-8.76	6.11

Table 5.4.7 (cont.)

	9	10	11	12
1	-56.00	58.70	138.90	6.70
2	-27.50	125.40	178.60	111.90
3	-.11	5.92	16.41	2.64
4	-7.50	11.80	17.26	4.54
5	8.45	-.33	20.70	8.14
6	8.59	6.33	4.21	5.69
7	6.46	9.90	24.11	-8.76
8	21.89	10.09	5.48	6.11
9	38.85	10.06	6.07	2.82
10	10.06	50.18	8.53	7.47
11	6.07	8.53	51.26	10.45
12	2.82	7.47	10.45	47.32

Table 5.4.8

Matrix E Middle (ages: 11 yr. to 12 yr. 11 mo.)

	1	2	3	4	5	6	7	8
1	11349.00	8922.00	331.40	258.90	421.00	548.20	584.50	- 3.90
2	8922.00	28755.00	678.30	590.90	648.10	831.90	974.80	- 71.10
3	331.40	678.30	87.57	75.59	89.60	70.13	67.20	- 7.08
4	258.90	590.90	75.59	76.93	85.11	60.12	63.73	- 1.29
5	421.00	648.10	89.60	85.11	147.26	82.37	85.50	5.41
6	548.20	831.90	70.13	60.12	82.37	95.74	84.63	- 2.38
7	584.50	974.80	67.20	63.73	85.50	84.63	148.19	- 6.59
8	- 3.90	- 71.10	- 7.08	- 1.29	5.41	- 2.38	- 6.59	23.75
9	134.40	69.70	5.15	1.38	12.56	12.05	10.80	9.32
10	47.10	202.50	11.11	11.51	14.39	16.70	14.14	3.49
11	45.90	12.90	10.44	9.80	14.55	- 1.02	6.47	4.22
12	38.30	11.00	2.80	.51	5.74	9.03	- 1.31	1.58

Table 5.4.8 (cont.)

	9	10	11	12
1	134.40	47.10	45.90	38.30
2	69.70	202.50	12.90	11.00
3	5.15	11.11	10.44	2.80
4	1.38	11.51	9.80	.51
5	12.56	14.39	14.55	5.74
6	12.05	16.70	- 1.02	9.03
7	10.80	14.14	6.47	- 1.31
8	9.32	3.49	4.22	1.58
9	17.83	1.60	3.63	3.27
10	1.60	29.96	- 1.18	2.62
11	3.63	- 1.18	30.12	2.79
12	3.27	2.62	2.79	20.06

Table 5.4.9

Matrix E Old (ages: 13 yr. up)

	1	2	3	4	5	6	7	8
1	31175.00	23126.00	967.90	1011.90	977.00	1567.10	1638.70	172.90
2	23126.00	32364.00	1055.00	1016.10	1023.10	1582.20	1658.90	53.30
3	967.90	1055.00	220.03	181.20	181.27	157.37	120.36	- 15.14
4	1011.90	1016.10	181.20	209.09	196.06	153.00	114.73	18.87
5	977.00	1023.10	181.27	196.06	300.48	170.97	145.57	23.34
6	1567.10	1582.20	157.37	153.00	170.97	213.15	188.88	15.50
7	1638.70	1658.90	120.36	114.73	145.57	188.88	257.46	13.58
8	172.90	53.30	- 15.14	18.87	23.34	15.50	13.58	74.77
9	48.90	7.70	19.46	- .48	21.29	15.76	16.16	10.27
10	- 1.40	29.60	14.62	13.94	8.84	8.04	5.16	9.95
11	25.40	- 149.20	5.23	2.63	12.25	- 8.89	9.33	8.83
12	10.00	91.40	14.63	9.35	8.21	6.88	- 6.44	1.26

Table 5.4.9 (cont.)

	9	10	11	12
1	48.90	- 1.40	25.40	10.00
2	7.70	29.60	-149.20	91.40
3	19.46	14.62	5.23	14.63
4	- .48	13.94	2.63	9.35
5	21.29	8.84	12.25	8.21
6	15.76	8.04	- 8.89	6.88
7	16.16	5.16	9.33	- 6.44
8	10.27	9.95	8.83	1.26
9	43.93	11.57	9.28	8.03
10	11.57	52.84	5.33	2.20
11	9.28	5.33	49.98	7.22
12	8.03	2.20	7.22	58.89

and special schools on all three age groups are shown in Table 5.4.10, and the difference vectors, $\bar{\underline{d}}$, are shown in Table 5.4.11.

The discriminant functions $\underline{a}'\underline{z}$, are as follows: (variables z_1 and z_2 in months, variables $z_3, z_4, \ldots, z_{12}$ in grade (year) equivalents. The values of the $\underline{a}$'s were multiplied, in each case, by a convenient power of 10.).

Young

Set I (variables 1, 2, 6, 7):

$$\underline{a}'\underline{z} = 3.32\ z_1 + .09\ z_2 - 50.80\ z_6 + 7.68\ z_7$$

Set II (variables 3, 4, 5, 6, 7):

$$\underline{a}'\underline{z} = 1.32\ z_3 - .31\ z_4 - .11\ z_5 - .15\ z_6 - .42\ z_7$$

Set III (variables 8, 9, 10, 11):

$$\underline{a}'\underline{z} = 1.50\ z_8 - .88\ z_9 - .90\ z_{10} - .67\ z_{11}$$

Middle

Set I :

$$\underline{a}'\underline{z} = .05\ z_1 - .54\ z_2 - 74.51\ z_6 + 46.52\ z_7$$

Set II:

$$\underline{a}'\underline{z} = .72\ z_3 + 4.02\ z_4 + 5.13\ z_5 + 1.62\ z_6 - 6.00\ z_7$$

Set III:

$$\underline{a}'\underline{z} = 3.08\ z_8 - 4.52\ z_9 + 9.43\ z_{10} + 4.38\ z_{11}$$

Old

Set I:

$$\underline{a}'\underline{z} = .88\ z_1 - 1.07\ z_2 - 52.96\ z_6 + 25.89\ z_7$$

Table 5.4.10*

Means (Public Schools)

Var	Young	Middle	Old
1	76.9500	95.5000	110.0556
2	76.8250	97.0250	112.2847
3	1.8975	2.4425	3.3292
4	1.8688	2.4825	3.2569
5	1.8638	2.8200	3.6389
6	1.4850	2.4475	3.5139
7	1.7762	2.7450	3.8569
8	.4750	.3025	.2944
9	.4262	.3925	.2625
10	.5088	.3025	.2806
11	.4850	.3375	.2833
12	.5488	.4225	.3583

Means (Special Schools)

Var	Young	Middle	Old
1	79.2417	93.7000	106.2685
2	77.1417	93.8333	106.7685
3	1.2525	1.9050	2.5917
4	1.4108	1.9667	2.5768
5	1.4083	2.0367	2.7954
6	1.3125	2.0850	2.8426
7	1.7325	2.7533	3.4898
8	.5092	.3850	.3463
9	.4117	.3717	.3944
10	.4642	.5550	.3231
11	.4458	.4550	.4426
12	.5483	.3683	.3981

*Mental Age in months, achievement measures in grade equivalents.

Table 5.4.11

Mean Differences (Public minus Special)

Var	Young	Middle	Old
1	-2.2917	1.8000	3.7870
2	- .3167	3.1917	5.5162
3	.6450	.5375	.7375
4	.4579	.5158	.6801
5	.4554	.7833	.8435
6	.1725	.3625	.6713
7	.0438	- .0083	.3671
8	- .0322	- .0825	- .0518
9	.0146	.0208	- .1319
10	.0446	- .2525	- .0426
11	.0392	- .1175	- .1592
12	.0004	.0542	- .0398

Set II:

$$\underline{a}'\underline{z} = 1.54\ z_3 - .70\ z_4 + 1.58\ z_5 + 3.25\ z_6 - 2.26\ z_7$$

Set III:

$$\underline{a}'\underline{z} = .04\ z_8 + 2.42\ z_9 - .01\ z_{10} + 2.73\ z_{11}\ .$$

The discriminant functions for each age group on all variables are:

Young

$$\begin{aligned}\underline{a}'\underline{z} = {} & 3.68\ z_1 + .08\ z_2 - 181.68\ z_3 + 63.98\ z_4 + 18.47\ z_5 \\ & + .75\ z_6 + 34.58\ z_7 - 62.35\ z_8 + 41.11\ z_9 - 10.94\ z_{10} \\ & - 4.44\ z_{11} + 14.63\ z_{12}\end{aligned}$$

Middle

$$\begin{aligned}\underline{a}'\underline{z} = {} & .05\ z_1 + .11\ z_2 - 3.85\ z_3 + 10.48\ z_4 + 6.48\ z_5 \\ & + .12\ z_6 - 6.52\ z_7 - 5.80\ z_8 + 5.40\ z_9 - 13.54\ z_{10} \\ & - 8.47\ z_{11} + 3.01\ z_{12}\end{aligned}$$

Old

$$\begin{aligned}\underline{a}'\underline{z} = {} & .36\ z_1 - .71\ z_2 - 40.78\ z_3 + 36.34\ z_4 - 25.71\ z_5 \\ & - 25.86\ z_6 + 20.87\ z_7 - 12.13\ z_8 + 58.15\ z_9 + 3.53\ z_{10} \\ & + 19.12\ z_{11} + 10.90\ z_{12}\ .\end{aligned}$$

4.4 Ordering

In order to make a representative selection we must first order the variables within each "common-factor" set as to their importance. This is done by ordering them according to their absolute correlation with the artificial variable $\underline{a}'\underline{z}$. These correlations are given by Equation (5.2.1), which states:

$$\underline{r}' = \frac{1}{\sqrt{\underline{a}' E \underline{a}}} \underline{a}' E D_{1/\sqrt{e_{ii}}} \quad .$$

Now, since $\underline{a}' E = \bar{\underline{d}}'$,

$$\underline{r}' = \frac{1}{\sqrt{\bar{\underline{d}}' \underline{a}}} \bar{\underline{d}}' D_{1/\sqrt{e_{ii}}} \quad . \tag{5.4.10}$$

The correlations are then as follows:

Young

Set I:

$$\begin{array}{lcccc} & 1 & 2 & 6 & 7 \\ \underline{r}' = & [-.44, & -.04, & .42, & .09] \end{array}$$

Set II:

$$\begin{array}{lccccc} & 3 & 4 & 5 & 6 & 7 \\ \underline{r}' = & [.73, & .51, & .39, & .21, & .04] \end{array}$$

Set III:

$$\begin{array}{lcccc} & 8 & 9 & 10 & 11 \\ \underline{r}' = & [-.44, & .21, & .56, & .48] \end{array}$$

All variables:

$$\underline{r}' = [-.19, -.02, .62, .43, .34, .18, .04, -.05, .03, .07, .06, .00]$$

Middle

Set I:

$$\begin{array}{lcccc} & 1 & 2 & 6 & 7 \\ \underline{r}' = & [.31, & .35, & .69, & -.01] \end{array}$$

Set II:

$$r' = \begin{array}{ccccc} 3 & 4 & 5 & 6 & 7 \\ [.68, & .70, & .76, & .44, & -.01] \end{array}$$

Set III:

$$r' = \begin{array}{cccc} 8 & 9 & 10 & 11 \\ [-.30, & .87, & -.85, & -.38] \end{array}$$

All variables:

r' = [.14, .16, .48, .49, .54, .31, -.01, -.14, .00, -.41, -.18, .10]

Old

Set I:

$$r' = \begin{array}{cccc} 1 & 2 & 6 & 7 \\ [.40, & .57, & .86, & .43] \end{array}$$

Set II:

$$r' = \begin{array}{ccccc} 3 & 4 & 5 & 6 & 7 \\ [.86, & .81, & .84, & .80, & .40] \end{array}$$

Set III:

$$r' = \begin{array}{cccc} 8 & 9 & 10 & 11 \\ [-.22, & -.72, & -.21, & -.82] \end{array}$$

All variables:

r' = [.30, .43, .70, .66, .69, .65, .32, -.08, -.28, -.08, -.32, -.07]

The orderings are:

Young

Set I : (1, 6, 7, 2)
Set II : (3, 4, 5, 6, 7)
Set III: (10, 11, 8, 9)
All : (3, 4, 5, 1, 6, 10, 11, 8, 7, 9, 2, 12)

Middle

Set I : (6, 2, 1. 7)
Set II : (5, 4, 3, 6, 7)
Set III: (9, 10, 11, 8)
All : (5, 4, 3, 10, 6, 11, 2, 1 or 8, 12, 7, 9)

Old

Set I : (6, 2, 7, 1)
Set II : (3, 5, 4, 6, 7)
Set III: (11, 9, 8, 10)
All : (3, 5, 4, 6, 2, 7 or 11, 1, 9, 8 or 10, 12) .

4.5 Selection

If we wish to choose the three tests that are closest to the best discriminator (discriminant function) between schools, the overall discriminant function would indicate the three achievement tests: Paragraph Meaning, Word Meaning and Spelling (variables 3, 4, and 5) for all three age groups. It should be noted that these three variables all belong to the same common-factor; more specifically, they are all verbal achievement measures. While these three doubtlessly show the greatest relative difference between the public and special school groups, a future study based on these three alone would completely disregard the gain measures and intelligence measures originally considered.

A representative selection (one variable from each factor) would lead to the following choice:

Young; Binet Mental Age, Achievement Paragraph Meaning, Gain Spelling (variables 1, 3, 10)

Middle; Arithmetic Reasoning, Achievement Spelling, Gain Word Meaning (variables 6, 5, 9)

Old; Arithmetic Reasoning, Achievement Paragraph Meaning, Gain Arithmetic Reasoning (variables 6, 3, 11).

It must be noted that, for the middle and old group, Arithmetic Reasoning showed up as the strongest discriminator on the Intelligence factor (even though the test was denoted as an "Achievement" test, and had high loadings on the achievement factor, too). The achievement factors in this study are probably more strongly determined by the verbal tests (see Table 5.4.6). The Arithmetic Achievement tests were as clearly present on the Intelligence as on the Achievement Factor; on the former, they even represent the best discriminator.

If additional experiments were to be performed with a reduced set of variables, and if one were interested in those variables only which show the strongest relative difference between the two groups, the three verbal achievement tests found by the overall discriminant analysis should be chosen. This subset would not, however, be representative of the whole set of original variables.

The choice of one of the representative sets (depending on age of subjects) would probably not lead to as strong a differentiation as the former, but all characteristics present in the original variables would be represented in the reduced set.

REFERENCES

[1] Aikin, Wilford M., "*The Story of the Eight-Year Study*", Harper and Broth., 1942.

[2] Anderson, T. W., "*An Introduction to Multivariate Statistical Analysis*", Wiley, 1958.

[3] Anderson, T. W., "Some scaling models and estimation procedures in the latent class model" in *Probability and Statistics* (Grenander, ed.), Wiley, 1959.

[4] Bargmann, Rolf, "A study of independence and dependence in multivariate normal analysis", *Inst. of Stat., North Carolina*, No. 186 (1957).

[5] Bargmann, Rolf E., and Mah, Valiant, "*Lecture Notes on the Theory of Least Squares*", Virginia Polytechnic Institute, 1961.

[6] Bartlett, M. S., "Tests of significance in factor analysis", *Brit. Journ. Psych.*, (Stat. Section) *1* (1948), 73-81.

[7] Dixon, W. J., and Massey, F. J., "*Introduction to Statistical Analysis*", McGraw Hill, 1957.

[8] Fisher, R. A., "The use of multiple measurements in taxonomic problems", *Ann. Eugen.*, *7* (1936), 179-186.

[9] Graybill, Franklin A., "*An Introduction to Linear Statistical Models*", McGraw Hill, 1961.

[10] Heck, D. L., "Charts of some upper percentage points of the distribution of the largest characteristic root", *Ann. Math. Stat.*, *31* (1960), 625-642.

[11] Hotelling, Harold, "Relations between two sets of variates", *Biometrika*, 28 (1936), 321-377.

[12] Howe, W. G., "Some contributions to factor analysis", *Oak Ridge Nat. Lab.*, ORNL 1919 (Physics), 1955.

[13] Kempthorne, Oscar, "*The Design and Analysis of Experiments*", Wiley, 1952.

[14] Kendall, M. G., "The Advanced Theory of Statistics", Vol. II, Griffin, 1951, 3rd ed.

[15] Kullback, S., "Information Theory and Statistics", Wiley, 1959.

[16] Monahan, Irene P., "Incomplete variable designs in multivariate experiments", Ph.D. dissertation, Virginia Polytechnic Institute, 1961.

[17] Pearson, E. S., and Hartley, H. O., "Biometrika Tables for Statisticians", Vol. I, Cambridge Univ. Press, 1956.

[18] Rao, C. Radhakrishna, "Advanced Statistical Methods in Biometric Research", Wiley, 1952.

[19] Roy, J., "Step-down procedure in multivariate analysis", Ann. Math. Stat., 29, (1958), 1177-1187.

[20] Roy, S. N., "Some Aspects of Multivariate Analysis", Wiley, 1957.

[21] Roy, S. N., and Bargmann, R. E., "Tests of multiple independence and the associated confidence bounds", Ann. Math. Stat., 29(1958), 491-503.

[22] Sokolnikoff, Ivan S., and Sokolnikoff, Elizabeth S., "Higher Mathematics for Engineers and Physicists", McGraw Hill, 1941 (2nd ed.).

[23] Thurstone, Louis L., "Multiple Factor Analysis", Univ. of Chicago Press, 1940.

[24] Thurstone, Thelma G., "An evaluation of educating Mentally handicapped children in special classes and regular classes", Coop. Research Grant No. 168(6452), U. S. Office of Education, Department of Health, Education, and Welfare, 1960.

REPRESENTATIVE ORDERING AND SELECTION OF VARIABLES

by

Rolf E. Bargmann

Part B
(Appendices)

Final Report (Part III, according to Proposal, Appendix A) of Cooperative Research Project No. 1132, "Statistical Models for the Evaluation and Interpretation of Educational Criteria", U. S. Office of Education, Department of Health, Education, and Welfare.

Blacksburg, Virginia

Virginia Polytechnic Institute

June, 1962

APPENDIX A

MULTIVARIATE ANALYSIS OF A STANDARD TWO-WAY CLASSIFICATION

METHODS OF MULTIVARIATE ANALYSIS

by Rolf Bargmann

Multivariate Analysis of a Standard Two-Way Classification

1. Condensation of Data

Table 1 is a typical record of data for a two-way classification with three response variables. Each row within a cell represents the result of measurement on one experimental unit, for each of the three responses. Examples for such responses would be

Y_1 = Pulse Rate
Y_2 = Systolic Blood Pressure
Y_3 = Diastolic Blood Pressure,

or

Y_1 = Score in an Achievement Test in English
Y_2 = Score in an Achievement Test in Arithmetic
Y_3 = Score in an Achievement Test in Science.

In both cases, the three measurements in a row in a given cell would come from the same person (experimental unit).

Examples for the A and B classification would be, for the first example,

A = Age: below 30, between 30 and 50, above 50.
B = Treatment, i.e., B_1 = Placebo B_2 = Treatment.

For the second example, the classification could be:

A: Intelligence Range: IQ 60 and below; IQ 60-70; IQ 70 and above.
B: School Type: B_1 = Special Schools for Retarded Children; B_2 = Public Schools.

(Actual data for each of these two examples were analyzed, but the data given here are artificial, for ease of computation.)

TABLE 1

		B					
		1			2		
A	1	40.1	85.6	46.8	63.0	102.4	44.8
		41.1	83.2	41.7	61.9	100.3	39.4
		40.9	79.5	38.1	61.6	101.3	39.9
		39.4	78.0	39.6	64.0	106.2	50.0
	2	52.8	114.4	47.1	68.6	129.8	42.7
		53.6	115.6	42.1	70.7	131.0	47.1
		53.9	114.2	42.3	69.1	135.8	45.2
		53.8	113.2	35.7	73.3	147.6	49.2
	3	41.6	142.3	45.8	62.8	164.4	44.4
		37.7	137.0	37.0	56.6	156.0	45.3
		43.2	143.8	44.0	63.3	161.3	46.3
		42.0	143.6	40.8	60.9	161.4	42.2

4 observation vectors per cell

3 random variables,

3 levels in A, 2 levels in B.

$\sum_{\text{all}} y_{j1}^2 = 75{,}171.35$

$\sum_{\text{all}} y_{j2}^2 = 379{,}031.57$

$\sum_{\text{all}} y_{j3}^2 = 45{,}175.55$

$\sum_{\text{all}} y_{j1} y_{j2} = 164{,}245.75$

$\sum_{\text{all}} y_{j1} y_{j3} = 57{,}328.77$

$\sum_{\text{all}} y_{j2} y_{j3} = 128{,}141.87$

Sum of squares and products for all 24 observations have been computed and recorded on the bottom of Table 1.

Just as in a univariate analysis, the data within each cell are combined into subtotals. In the multivariate case, however, this summing over the four observations in each cell must be done for each variable separately. To the right and on the bottom of the table are the marginal totals, again as in the univariate case, for each of the three variables separately.

TABLE 2
Subtotals

161.5	326.3	166.2	250.5	410.2	174.1	412.0	736.5	340.3
214.1	457.4	167.2	281.7	544.2	184.2	495.8	1001.6	351.4
164.5	566.7	167.6	243.6	643.1	178.2	408.1	1209.8	345.8
540.1	1350.4	501.0	775.8	1597.5	536.5	1315.9	2947.9	1037.5

At this stage it is very important to check the arithmetic. For this purpose a table has been prepared which consists of the sum of all measurements for each experimental unit, i.e., the three elements in each row in a cell have been added together. One obtains the table of subtotals and the sum of squares of all observations for these sums, in the customary manner for a univariate analysis.

TABLE 3
Sums-across-Variables
(Check Table)

	1	2
1	172.5 166.0 158.5 157.0	210.2 201.6 202.8 220.2
2	214.3 211.3 210.4 202.7	241.1 248.8 250.1 270.1
3	229.7 211.7 231.0 226.4	271.6 257.9 270.9 264.5

$$\sum_{\text{all}}\sum y^2_{jr} = 1,198,811.25$$

Subtotals

654.0	834.8	1488.8
838.7	1010.1	1848.8
898.8	1064.9	1963.7
2391.5	2909.8	5301.3

Now the following checks are available:

$$\sum_{\text{all}}\sum y^2{}_{jr} = \sum_{\text{all}} y^2{}_{j1} + \sum_{\text{all}} y^2{}_{j2} + \sum_{\text{all}} y^2{}_{j3}$$

$$+ 2 \left[\sum_{\text{all}} y_{j1} \, y_{j2} + \sum_{\text{all}} y_{j1} \, y_{j3} + \sum_{\text{all}} y_{j2} \, y_{j3} \right]$$

Also, the sum of each row in each cell of Table 2 must be equal to the entries in the Subtotal Table in Table 3. Of course, to allow checking, the Subtotals Table in Table 3 should have been computed from the Main Table in Table 3.

With all these checks completed we can now proceed to estimate main effects, in the customary way, for each of the three variables separately. All estimates pertaining to variable Y_1 have a superscript (1). Then , by the familiar relations

$$\hat{\mu}^{(1)} = \sum_{\text{all}} y_{j1} \ / \ 24 = 54.8292$$

$$\hat{\alpha}_1^{(1)} = Y_{1..}^{(1)} \ / \ 8 - \hat{\mu}^{(1)}$$

$$= 412.0/8 - 54.8292$$

$$= -3.3292,$$

from that column of the right-hand margin of Table 2 which pertains to variable 1. The symbol $Y_{1..}^{(1)}$ denotes the sum overall observations in a cell (last dot) and over all observations in the first row (first dot) on variable 1 (superscript).
By the same argument: $\hat{\alpha}_2^{(1)} = Y_{2..}^{(1)} \ / \ 8 - \hat{\mu}^{(1)}$,
and $\hat{\alpha}_3^{(1)} = Y_{3..}^{(1)} / \ 8 - \hat{\mu}^{(1)}$. The values $Y_{1..}^{(1)}$, $Y_{2..}^{(1)}$, and $Y_{3..}^{(1)}$ are, of course, found in the first column (variable 1) of the right-hand margin of Table 2.

We can now do the same for the elements in the first column of the lower margins, in order to obtain:

$$\hat{\beta}_1^{(1)} = Y_{.1.}^{(1)} \ / \ 12 - \hat{\mu}^{(1)} = 540.1/12 - 54.8292$$

$= -9.8209$, and the same can be done for $\hat{\beta}_2^{(1)}$. Notice the divisor 8 in the estimation of α (where 8 observations make up the total) and the divisor 12 in the estimation of β (where 12 observations make up the total). The same procedure has been followed for variables 2 and 3, based upon the second and third columns within each cell of Table 2. The results are:

TABLE 4: Estimates of Effects

$\hat{\mu}^{(1)} = 54.8292$	$\hat{\mu}^{(2)} = 122.8292$	$\hat{\mu}^{(3)} = 43.2292$
$\hat{\alpha}_1^{(1)} = -3.3292$	$\hat{\alpha}_1^{(2)} = -30.7667$	$\hat{\alpha}_1^{(3)} = -0.6917$
$\hat{\alpha}_2^{(1)} = 7.1458$	$\hat{\alpha}_2^{(2)} = 2.3708$	$\hat{\alpha}_2^{(3)} = 0.6958$
$\hat{\alpha}_3^{(1)} = -3.8167$	$\hat{\alpha}_3^{(2)} = 28.3958$	$\hat{\alpha}_3^{(3)} = -0.0042$
$\hat{\beta}_1^{(1)} = -9.8209$	$\hat{\beta}_1^{(2)} = -10.2959$	$\hat{\beta}_1^{(3)} = -1.4792$
$\hat{\beta}_2^{(1)} = 9.8208$	$\hat{\beta}_2^{(2)} = 10.2958$	$\hat{\beta}_2^{(3)} = 1.4791$

A check is available on these computations in that sums over α_i and β_j must be zero, except for rounding. This is due to the usual constraints in a two-way table with equal numbers of replications per cell. If the numbers of replications per cell are different but proportional, the same analysis as here described can be performed, but now the weighted sums add to zero, where the weights are the number of observations having the same effect. If the numbers of replications are disproportionate, the analysis is more complicated and must be performed by extending the general two-way classification analysis to the multivariate case, in a manner completely analogous to the one here described.

Now, if we denote by G_1 the grand total of all observations on the first variable (1315.9 from Table 2), by G_2 that on the second and by G_3 that on the third, we must have the following "correction terms":

$G_1^2 / 24 = 72,149.70$ $\quad$ $G_2^2 / 24 = 362,088.10$

$G_3^2 / 24 = 44,850.26$ $\quad$ $G_1 G_2 / 24 = 161,630.90$

$G_1 G_3 / 24 = 56,885.26$ $\quad$ $G_2 G_3 / 24 = 127,435.26$

For a check, we obtain the correction term from Table 3, i.e.,

$(5,301.3)^2 / 24 = 1,170,990.90$,

and this must equal the sum of the "square" terms above plus twice the sum of the "cross-product" terms.

We can now compute the Analysis-of-Variance Table (Table 5).

The reader who is familiar with the analysis-of-covariance with several concomitant variables will need no explanation on how to compute the elements of this table. The columns headed Y_1^2, Y_2^2, Y_3^2 are simply the standard analysis-of-variance expressions on each of the three variables separately. The cross-product terms are obtained by direct analogy, e.g.,

Sum of Products due to A in the $Y_1 Y_2$ column

$$= Y_{1..}^{(1)} Y_{1..}^{(2)} / 8 + Y_{2..}^{(1)} Y_{2..}^{(2)} / 8 + Y_{3..}^{(1)} Y_{3..}^{(2)} / 8$$

$$- G_1 G_2 / 24 \;;$$

Sum of Products of Subtotals in the $Y_1 Y_2$ column

$$= Y_{11.}^{(1)} Y_{11.}^{(2)} / 4 + Y_{12.}^{(1)} Y_{12.}^{(2)} / 4 + \ldots + Y_{32.}^{(1)} Y_{32.}^{(2)} / 4$$

$-G_1 G_2 / 24;$ (elements in the body of Table 2).

Sum of Products (Total) in the $Y_1 Y_2$ column

$$= \sum_{\text{all}} y_{j1} y_{j2} - G_1 G_2 / 24.$$

The error and A X B rows are obtained in the customary manner by forming differences. Notice that the cross-product terms may be negative, but the square terms must be positive.

TABLE 5 Analysis of Variance

Source	df*	Y_1^2	Y_1Y_2	Y_1Y_3	Y_2^2	Y_2Y_3	Y_3^2	(Check) $(Y_1+Y_2+Y_3)^2$
A	2	613.71	87.93	58.33	14,068.26	182.49	7.70	15,347.17
B	1	2,314.77	2,426.73	348.64	2,544.10	365.50	52.51	11,193.12
A x B	2	28.67	-4.47	-12.30	7.20	3.04	5.46	13.86
Subtotal	5	2,957.15	2,510.19	394.67	16,619.56	551.03	65.67	26,554.15
Error	18	64.50	104.66	48.84	323.91	155.58	259.62	1,266.20
Total	23	3,021.65	2,614.85	443.51	16,943.47	706.61	325.29	27,820.35

* univariate degrees of freedom.

The last column is a check on the arithmetic. It represents the standard analysis-of-variance table based upon the data in Table 3. In each row of Table 5, the sum of the elements in the "square" columns plus twice the sum of the elements in the "cross-product" columns must add to the "check" column entry.

2. Estimation of Covariances and Correlations

The elements in the "Error" row of Table 5 can now be collected into an "error Matrix" E, which is the multivariate analogue of a "sum-of-squares due to error". The "square" terms appear in the diagonals and the "cross-product" terms in the off-diagonals of the symmetric matrix E.

Table 6

E

$$\begin{bmatrix} 64.50 & 104.66 & 48.84 \\ 104.66 & 323.91 & 155.58 \\ 48.84 & 155.58 & 259.62 \end{bmatrix}$$

If desired, an unbiased estimate of Σ can now be computed by division of each element in E by the (univariate) degrees of freedom due to error. This df is always denoted by the symbol

n_e = d.f. (error),

in our case, $n_e = 18$ (see Table 5).

Table 7

$S = 1/18\ E$

$$\begin{bmatrix} 3.5833 & 5.8144 & 2.7133 \\ 5.8144 & 17.9950 & 8.6433 \\ 2.7133 & 8.6433 & 14.4233 \end{bmatrix}$$

If desired, the matrix of correlations between the three variables can be computed. For this purpose we compute the square root of the diagonal elements of E.

$\sqrt{e_{11}} = 8.031,189 \quad \sqrt{e_{22}} = 17.997,501 \quad \sqrt{e_{33}} = 16.112,728,$

and divide each off-diagonal element e_{ij} by $\sqrt{e_{ii}}\,\sqrt{e_{jj}}$. This yields the estimates of the correlation coefficients

Table 8

R

$$\begin{bmatrix} 1 & .7241 & .3774 \\ .7241 & 1 & .5365 \\ .3774 & .5365 & 1 \end{bmatrix}$$

At this stage we must make a decision on the order in which we wish to attack the three variables, for purposes of a "step-down" analysis to be performed later. If the experimenter specifies the order, in terms of what he considers to be the order of importance for the physical variables, the statistician should not interfere, but adopt the experimenter's ordering principle. If the experimenter does not specify a hierarchial order, the statistician may adopt one on the basis of the data available; he could change it if that were desirable in the further progress of the analysis.

An inspection of Table 5 shows that, for all effects, variable 2 has the largest sums-of-squares (column Y_2^2), and variable 3 has the smallest (column Y_3^2). It would thus seem appropriate to adopt the order

2-1-3.

From here on, this order will be adopted in all our analyses. We must thus rewrite the E matrix to conform to this order.

Table 9
E (in order 2-1-3)

$$\begin{vmatrix} 323.91 & 104.66 & 155.58 \\ 104.66 & 64.50 & 48.84 \\ 155.58 & 48.84 & 259.62 \end{vmatrix}$$

In the sequel we will need, repeatedly,
E^{-1}, the triangular matrix $\tilde{T}$ such that
$\tilde{T}T' = E$, the matrix $\tilde{T}^{-1}$, $|E|$,
and the "leading coefficients" of E. It is thus advisable to perform inversion by the Doolittle method which will give us all these expressions at once. Actually we will invert $10^{-2}E$, to keep numbers in manageable form.

Table 10
Doolittle Inversion of $10^{-2}E$, order 2-1-3

						Σ	ch
3.2391	1.0466	1.5558	1	0	0	6.4815	
1	.32311	.48032	.30873	0	0	2.11216	2.11216
	.6450	.4884	0	1	0	2.1334	
	.30683	-.01429	-.32311	1	0	.96943	.96943
	1	-.04657	-1.05306	3.25913	0	3.15950	3.15950
		2.5962	0	0	1	3.5962	
		1.84825	-.49537	.04657	1	2.39945	2.39945
		1	-.26802	.02520	.54105	1.29823	1.29823

$$\begin{bmatrix} .78175 & -1.06554 & -.26802 \\ -1.06554 & 3.26030 & .02520 \\ -.26802 & .02520 & .54105 \end{bmatrix} \times 10^{-2} = E^{-1}$$

The three underscored elements, 3.2391, .30683, and 1.84825 are the "leading" coefficients. We need the square-roots of these quantities, for later use in obtaining triangular matrices. They are

$$\sqrt{\ell_{11}} = 1.79975 \quad \sqrt{\ell_{22}} = .55392 \quad \sqrt{\ell_{33}} = 1.35950$$

Notice that these are the leading coefficients for $10^{-2}E$. They must be multiplied by 100 in order to become the leading coefficients of E, and the square-roots must be multiplied by 10.

3. Union-Intersection Tests and Discriminant Functions

As in an ordinary analysis of variance, the interaction test should be applied first for the purpose of determining whether the two-way classification should be retained or abandoned. For every hypothesis to be tested the following expressions must be evaluated:

$$s = \min [n_h, p]$$
$$m = (|n_h - p| - 1)/2$$
$$n = (n_e - p - 1)/2$$

where n_h = d.f. due to a given hypothesis,

n_e = d.f. due to error

p = number of variables

The parameters s, m, and n are the ones used in Heck's charts (D. L. Heck, "Charts of some upper percentage points of the distribution of the largest characteristic root", Ann. Math. Stat. vol. 31 (1960) pp. 625-642).

3.1 Interaction A x B

From the row (A x B) in Table 5 we obtain a matrix called H (the multivariate analogue of a "sum-of-squares due to hypothesis"). We write it in the order 2-1-3 to make it conform to the ordering of variables.

Table 11

$$\begin{array}{c} H \\ \begin{bmatrix} 7.20 & -4.47 & 3.04 \\ -4.47 & 28.67 & -12.30 \\ 3.04 & -12.30 & 5.46 \end{bmatrix} \end{array}$$

To test for the presence of interaction, we need the largest characteristic root of $E^{-1}H$. However, since $n_h = 2$, H can be at most of rank 2, and we may thus form $H = TT'$, where T is a (3 x 2) matrix. Then $ch_{max}(E^{-1}H) = ch_{max}(T'E^{-1}T)$ and the matrix in brackets is of order (2 x 2) and symmetric. The matrix T can be obtained by the forward Doolittle method applied to H.

$$\begin{array}{ccc}
\underline{7.20} & 4.47 & 3.04 \\
 & -.6208 & .4222 \\ \hline
 & 28.67 & -12.30 \\
 & \underline{25.89560} & -10.4128 \\
 & & -.4021 \\ \cline{2-3}
 & & 5.46 \\
 & & -.01
\end{array}$$

The last leading coefficient is zero within the precision of the elements of H. T' is obtained by dividing each row starting with a leading coefficient by the square-root of that leading coefficient. Hence

$$T' = \begin{bmatrix} 2.68 & -1.67 & 1.13 \\ 0 & 5.09 & -2.05 \end{bmatrix}$$

thus

$$T'E^{-1} = 10^{-2} \begin{bmatrix} 2.68 & -1.67 & 1.13 \\ 0 & 5.09 & -2.05 \end{bmatrix} \begin{bmatrix} .782 & -1.065 & -.268 \\ -1.065 & 3.260 & .025 \\ -.268 & .025 & .541 \end{bmatrix} \begin{matrix} (-.551) \\ (2.220) \\ (\ .298) \end{matrix}$$

$$= 10^{-2} \begin{bmatrix} 3.571 & -8.270 & -.149 \\ -4.871 & 16.542 & -.982 \end{bmatrix} \begin{matrix} (-4.847) \\ (10.689) \end{matrix}$$

The numbers in parentheses after the second matrix are the sums of rows in the second matrix. The numbers in parentheses after the resulting matrix are the values obtained by multiplying T' by the sum column above and must, within rounding, equal the sums of rows in the result matrix.

$$T'E^{-1}T = 10^{-2} \begin{bmatrix} 3.571 & -8.270 & -.149 \\ -4.871 & 16.542 & -.982 \end{bmatrix} \begin{bmatrix} 2.68 & 0 \\ -1.67 & 5.09 \\ 1.13 & -2.05 \end{bmatrix} \begin{pmatrix} 2.68 \\ 3.42 \\ -.92 \end{pmatrix}$$

$$= 10^{-2} \begin{bmatrix} 23.21 & -41.79 \\ -41.79 & 86.21 \end{bmatrix} \begin{pmatrix} -18.58 \\ 44.42 \end{pmatrix}$$

Hence $T'E^{-1}T = \begin{bmatrix} .2321 & -.4179 \\ -.4179 & .8621 \end{bmatrix}$

The equation which produces the characteristic roots of this matrix is

$\lambda^2 - \text{tr}\,(T'E^{-1}T)\lambda + |T'E^{-1}T| = 0$, or

$\lambda^2 - 1.0942\,\lambda + .025{,}453 = 0$

$\lambda = .5471 \pm .5233$

$\lambda\ \max = 1.0704 = c$, say.

To reduce this to standard form we must compute

$\theta = \frac{c}{1+c} = .517.$

The parameters for Heck's charts are

$s = 2;\ m = (\ |2-3|\ -1)/2 = 0$

$n = (18-3-1)/2 = 7.$ From Heck's charts,

$\theta_{.05} = .470$ and $\theta_{.01} = .577$

Since $.517 < .577$, we accept the null hypothesis of no interaction effects at the .01 level but not at the .05 level. There may be some doubt whether we should abandon the two-way classification at this stage. For a further discussion, see page 28.

3.2 A-Effect

H, for A-effect, in the order 2-1-3 for the variables, is

Table 12

$$10^2 \begin{bmatrix} & H & \\ 140.6826 & .8793 & 1.8249 \\ .8793 & 6.1371 & .5833 \\ 1.8249 & .5833 & .0770 \end{bmatrix}$$

Again, $n_h = 2$, so that we can reduce the matrix $E^{-1}H$ to one of order (2 x 2) for the purpose of finding the largest root.

Forward Doolittle on $10^{-2}H$:

$$\begin{array}{lll} 140.6326 & .8793 & 1.8249 \\ 1 & .006250 & .012972 \\ \hline & 6.1371 & .5833 \\ & 6.131604 & .571894 \\ & 1 & .093270 \\ \hline & & .0770 \\ & & -.00001 \end{array}$$

$$T' = 10 \begin{bmatrix} 11.8610 & .0741 & .1539 \\ 0 & 2.4762 & .2310 \end{bmatrix}$$

$$T'E^{-1}T = \begin{bmatrix} 107.1658 & -31.4007 \\ -31.4008 & 20.0485 \end{bmatrix}$$

$$\lambda^2 - 127.215\lambda + 1162.54833 = 0$$

$$\lambda_{max} = c = 117.3045$$

$$\theta = c/(1 + c) = .9915$$

$s = 2$, $m = 0$, $n = 7$. $\theta_{.01} = .577$.

$$.9915 \gg .577,$$

hence the Λ effect is very significant.

To obtain the discriminant function among the variables which produces maximum separation of the Λ-effects we must solve

$$E^{-1}H\,\underline{a} = \lambda\,\underline{a},$$

where λ is the largest root just obtained. We will, however, solve more easily,

$$T'E^{-1}T\underline{u} = \lambda\,\underline{u}, \text{ for then}$$

$$E^{-1}TT'E^{-1}T\underline{u} = \lambda E^{-1}T\underline{u}$$

$$\text{and hence } E^{-1}T\underline{u} = \underline{a}.$$

Now, since an eigenvector is unspecified in length, we may set the last element of $\underline{u}$ (here u_2) equal to 1. This is equivalent to disregarding the last row of $T'E^{-1}T$ and transposing the last column (with sign change) to the right-hand side.

For our case, then

$$(107.1658 - 117.3045)\,u_1 = 31.4007$$

$$u_1 = -3.09711$$

$$E^{-1}T\underline{u} = \begin{bmatrix} .9153 & -.2700 \\ -1.2392 & .8079 \\ -.3094 & .0187 \end{bmatrix} \begin{bmatrix} -3.09711 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -3.1048 \\ 4.6458 \\ .9769 \end{bmatrix}$$

The matrix $E^{-1}T$ was available as an intermediate step in the calculation of $T'E^{-1}T$. The discriminant function which maximises the A-effect difference is thus, except for an arbitrary multiplier,

$$4.65Y_1 - 3.10Y_2 + 0.98Y_3 \;.$$

If we multiply the three values for each experimental unit by these weights, we obtain a composite score for which the F-ratio due to A will be as large as possible. Notice that the vector $\underline{a}$ is in the variable order 2-1-3, so that weights must be assigned to the proper variables.

3.3 B - Effect Order (2-1-1)

Since there is only one degree of freedom due to hypothesis, we proceed directly to an F test. The F statistic for H_0: $\hat{\underline{a}}_1 = \hat{\underline{a}}_2$

$$F = \frac{1}{p} \times \frac{v}{2} \times (n_e - p + 1)\ (\hat{\underline{a}}_2' - \hat{\underline{a}}_1')\ E^{-1}(\hat{\underline{a}}_2 - \hat{\underline{a}}_1),$$

and the weight of the discriminant function, $\hat{\underline{a}}$, is proportional to $E^{-1}(\hat{\underline{a}}_2 - \hat{\underline{a}}_1)$, where v is the number of observations on which each estimate is based, here 12, and $\hat{\underline{a}}_1'$, $\hat{\underline{a}}_2'$ are the vectors of estimates of β - effects, i.e., the row vectors in the last two lines of Table 4.

$$E^{-1}(\hat{\underline{a}}_2-\hat{\underline{a}}_1)=10^{-2} \begin{bmatrix} .7818 & -1.0655 & -.2680 \\ -1.0655 & 3.2603 & .0252 \\ -.2680 & .0252 & .5411 \end{bmatrix} \begin{bmatrix} 20.5917 \\ 19.6417 \\ 2.9583 \end{bmatrix}$$

$$= 10^{-2} \begin{bmatrix} -5.6225 \\ 42.1719 \\ -3.4229 \end{bmatrix}$$

Hence a discriminant function which produces maximum separation between the two groups denoted by B_1 and B_2 is

$$42Y_1 - 5.6Y_2 - 3.4Y_3.$$

It will be noted that, even though variable 2 has the greatest mean difference, the discriminant function places maximum weight on variable 1. This is partly due to the high correlation between variables 1 and 2.

$$F = \frac{1}{3} \times 6 \times 16 \times 10^{-2}$$

$$\times [-5.6225 \times 20.5917 + 42.1719 \times 19.6417 - 3.4229 \times 2.9583]$$

$$= 224.8 .$$

We must compare this with $[p,\ n_e - p + 1]$,

i. e., [3, 16] d.f.

$F_{.001}(3,16) = 127$. $224.8 > 127$, hence the B-effect is very significant.

An unbiased estimate of Δ^2, the square of the Mahalanobis distance, for a difference of effects is

$$p \left[\frac{1}{n_1} + \frac{1}{n_2} \right] \left[\frac{n_e-p-1}{n_e-p+1} F - 1 \right] = D^2$$

here $D^2 = 3 \times (1/12 + 1/12) \left[\frac{11}{16} \times 224.8 - 1 \right]$

$$= 97.85$$

Hence the two groups designated by B_1 and B_2 are almost 10 standard units apart.

3.4 All Effects (Subtotals) Order 2-1-3

In this case, H is of full rank, and we may compute $E^{-1}H$. However, since $E = \tilde{T}\tilde{T}'$, and hence $E^{-1} = (\tilde{T}^{-1})' \tilde{T}^{-1}$, and since $\tilde{T}^{-1}$ is readily obtained from the right-hand side of Table 10, we will compute

$\tilde{T}^{-1}H(\tilde{T}^{-1})'$, a symmetric matrix whose characteristic roots are the same as those of $E^{-1}H$. The eigenvector associated with the largest root is given by the relation

$$E^{-1}H\underline{a} = \lambda\underline{a}$$

and this is the weight function for the discriminant function. If we find the eigenvector $\underline{u}$ of the symmetric matrix above, i.e., if

$$\tilde{T}^{-1}H(\tilde{T}^{-1})'\underline{u} = \lambda\underline{u},$$

then $(\tilde{T}^{-1})' \tilde{T}^{-1}H(\tilde{T}^{-1})'\underline{u} = \lambda(\tilde{T}^{-1})'\underline{u}$

or $E^{-1}H\underline{a} = \lambda\underline{a}$,

where $\underline{a} = (\tilde{T}^{-1})'\underline{u}$.

Dividing the elements in the rows headed by the "leading coefficients" to the right of the first vertical line by the square-root of the leading coefficients we obtain.

$$\tilde{T}^{-1} = 10^{-1} \begin{bmatrix} .55563 & 0 & 0 \\ -.58332 & 1.80531 & 0 \\ -.36438 & .03426 & .73556 \end{bmatrix}$$

H in the order (2-1-3) is obtained from the Subtotal row of

Table 5, and is

$$H = 10^2 \begin{bmatrix} 165.1956 & 25.1019 & 5.5103 \\ 25.1019 & 29.5715 & 3.9467 \\ 5.5103 & 3.9467 & .6567 \end{bmatrix}$$

Hence $\tilde{T}^{-1} H (\tilde{T}^{-1})'$

$$= \begin{bmatrix} 51.3087 & -28.6863 & -30.9181 \\ -28.6863 & 100.0596 & 23.0163 \\ -30.9181 & 23.0163 & 19.0747 \end{bmatrix}$$

We will call this matrix F, and obtain the largest root and the eigenvector of 1/100 F by an iterative procedure, in order to illustrate the method.

a) Initialising

An important problem in finding the largest root and associated eigenvector by an iterative procedure is to find a good initial value of the elements of the so-called "principal components" (eigenvectors of such length that the sum of squares of elements equals the root). A very fast and relatively close guess can be obtained by the "Centroid" method, which proceeds as follows:

Table 13

Centroid Method on 1/100 F

				Σ
		- .28686	-.30918	-.59604
	- .28686		.23016	-.05670
	- .30918	.23016		-.07902
Σ	- .59604	- .05670	-.07902	-.73176
-1/2Σ	.29802	.02835	.03951	.36588
B (1)	(.29802)	- .25851	-.26967	-.23016
-2 B	- .59604	.51702	.53934	.46032
± d	- .51309	1.00060	.19075	.67826
C	-1.10913	1.51762	.73009	1.13858

$\Sigma\ |C| = 3.35684 \qquad \sqrt{\Sigma\ |C|} = 1.83217$

a_o	- .6054	.8283	.3985

Write the matrix omitting diagonal elements (Table 13). Sum rows and columns. (The right-hand column is carried as a check). Take -1/2 times the sum row. Select the largest positive element in this row. It is .29802 and occurs in variable 1. Since variable 1 has this largest element, the first row of the matrix is added to the row -1/2 Σ. Put the first (unchanged) element in parentheses and write a (1) in front of this row to indicate that variable (1) has been "reflected". Again, in this last row, select the largest positive element which is not in parentheses. (Parentheses, once introduced, must be retained throughout). If all elements without parentheses are negative, and all elements in parentheses positive, stop and write a B in front of the final row. It will be noted that the stage B has been reached here right away. (For more detail see L. L. Thurstone, Multiple Factor Analysis, University of Chicago Press, 1940).

Multiply the B row by -2. Now write the diagonal terms of the matrix with the sign equal to that of the -2B row. Add the last two rows to obtain the C - row. Form the sum of the absolute values in the C-row, extract the square root, and divide each element of C by this root. The resulting row is a_o, the first guess of the principal component associated with the largest root.

b) Iteration

One now performs the usual iteration by multiplying the matrix by the vector a_o and repeating. One might accelerate convergence by squaring the matrix but this method, so strongly recommended by theoreticians is, computationally, impractical, the reason being that the work in squaring a matrix is fully equivalent to 6 iterations, and that is about all that is required if one starts with a close initial guess.

Table 14

Principal Component Iteration on 1/100 F

	.51309	- .28686	- .30918
	- .28686	1.00060	.23016
	- .30918	.23016	.19075
a_0	- .6054	.8283	.3985
$a'F$	- .67144	1.07952	.46305
$a'a$	1.21139	1.15213	1.34402
a_1	- .5543	.9370	.3445
$a'F$	- .65971	1.16197	.43565
$a'a$	1.30390	1.25258	1.23524
a_2	- .5059	.9277	.3527
a_3	- .5115	.9274	.3522
a_4	- .51164	.92768	.35212
a_5	- .51149	.92762	.35215
a_6	- .51152	.92763	.35214
a_7	- .51152	.92763	.35214

$a_7' a_7 = 1.246,153$

The steps are as follows (Table 14):

a) Write the full matrix.

b) Obtain sum of squares of elements in a_0. Record in the first element of $a' a$. (Leave $a'F$ blank for the time being).

c) Multiply first row of the matrix by a_0, i.e., $-.51309 \times .6054 - .28686 \times .8283 - .30918 \times .3985 = -.67144$. Record result in first element of $a'F$.

d) Divide element in $a'F$ by element in $a'a$ and thus obtain first element in a_1.

e) Cross out first element in $\underline{a}_0$, it is henceforth to be replaced by the first element in $\underline{a}_1$. This illustrates how the principal component can be iterated by a Gauss-Seydel scheme, which accelerates convergence very considerably, especially for large matrices.

f) Compute $a_{11}^2 + a_{02}^2 + a_{03}^2$

$= .5543^2 + .8283^2 + .3985^2 = 1.15213.$

Record this result in the second element of row $\underline{a}'\underline{a}$ (above line $\underline{a}_1$).

g) Multiply second row by a_{11}, a_{02}, a_{03}; thus obtain second element in $\underline{a}'F$ row, 1.07952.

h) Divide this by corresponding element in $\underline{a}'\underline{a}$. This yields the second element in $\underline{a}_1$.

i. Cross out a_{02}, just as in step e.

j. Compute $a_{11}^2 + a_{12}^2 + a_{03}^2$

$= .5543^2 + .9370^2 + .3985^2 = 1.34402.$

Record this in the third element of $\underline{a}'\underline{a}$ above line $\underline{a}_1$.

This process is continued. An improved term is used for the matrix by vector multiplication as soon as it is available. Table 14 shows that the fifth place is established after 7 iterations. In practice, the larger the matrix the fewer complete iterations are required if the initial guess is obtained by the centroid technique.

A minimum of 3 or 4 iterations is, however, necessary in every case, so that for matrices larger than (30 x 30) the work may become tedious.

We now have

$$\lambda \max (\hat{T}^{-1} H(\hat{T}^{-1})') = 124.615,$$

and an associated eigenvector

$$\underline{u}' [-.51152, .92763, .35214].$$

The characteristic root equals the final sum of squares of elements (here multiplied by 100 since we performed iteration on 1/100 F.) The eigenvector remains unchanged, for it can be multiplied by an arbitrary scalar. The root is of course the same as that of $E^{-1}H$.

For the discriminant function we must evaluate

$$\underline{a} = (\hat{T}^{-1})' \underline{u}$$

$$\underline{a} = 10^{-1} \begin{bmatrix} .55563 & -.58332 & -.36438 \\ 0 & 1.80531 & .03426 \\ 0 & 0 & .73556 \end{bmatrix} \begin{bmatrix} -.51152 \\ .92763 \\ .35214 \end{bmatrix}$$

$$= 10^{-1} \begin{bmatrix} -.9536 \\ 1.6867 \\ .2590 \end{bmatrix}$$

Hence the discriminant function, which maximises the F ratio for all effects (Subtotals) is

$$16.9\, Y_1 - 9.5 Y_2 + 2.6 Y_3$$

Notice, again, the exchange or order, since the H and E matrices were recorded in variable order 2-1-3.

The test statistic is again

$$c = 124.615$$

$$\theta = c/(1 + c) = .99204$$

$$s = 3 \quad m = (|5 - 3| - 1)/2 = 1/2$$

$$n = (18 - 3 - 1)/2 = 7$$

$\theta_{.01} = .70$ (interpolated) from Heck's charts.

$.99204 \gg .70$, hence over-all contribution of effects is very significant, as expected. The test described in this section (3.4) would actually be performed first of all. It was mentioned here in the end since it was used for the purpose of illustrating numerical techniques for the evaluation of roots and eigenvectors of large matrices. (For a 3 x 3 matrix one would not actually use this method, but a method like the one here described is indispensable if the matrix is of order 7 x 7 or larger).

4. Likelihood - Ratio Tests and Step-Down Analysis

4.1 General Methods

The likelihood-ratio test, which leads to different test statistics and slightly different probability statements (unless $n_h = 1$), is based on the same matrices H and E described in the previous sections. The test statistic, called "Wilks' lambda statistic" is

$$\Lambda = |E| \,/\, |H + E| \;.$$

The distribution of this statistic, except for a very small term (order m^{-4} where m is determined by the sample size) is under the null hypothesis:

$$\begin{aligned} &\Pr [\, -m \log_e \Lambda > c \,] \\ &= \Pr [\, \chi^2_f > c \,] \\ &+ \delta/m^2 \, (\Pr[\chi^2_{f+4} > c] - \Pr[\chi^2_f > c]) \end{aligned}$$

where $m = n_e + \frac{1}{2}(n_h - p - 1)$

$f = n_h \cdot p$,

and $\delta = n_h \cdot p\,(p^2 + n_h^2 - 5)/48$

As before, n_h is the usual (univariate) expression for the d.f. due to hypothesis and n_e is d.f. error. The χ^2 - probabilities can be read from Table 7 of the Biometrika Tables for Statisticians, vol.1.

Calculations are usually made in terms of logs to the base 10 and eventually multiplied by 2.302,585.

The Step-Down technique, which depends on the ordering of variables, determines how much of the significance of effect difference is attributed to the first variable, how much does the second contribute over and beyond the first, the third over and beyond the first and second, and so on. In chapter 2 (discussion following Table 8) we adopted the order 2-1-3 for our demonstration study. The test-statistics employed here are the leading coefficients of the E matrix in the forward Doolittle scheme, and those of each (H + E) matrix. Under the null hypothesis the ratio

(leading coefficient of E) / (leading coefficient of H + E)

has the Incomplete Beta Distribution with parameters

$$a = (n_e - p + 1)/2, \ b = p/2$$

for the first leading coefficient ratio (the (1,1) element of the matrices themselves),

$$a = (n_e - p)/2, \ b = p/2$$

for the second leading coefficient ratio, and so on, subtracting 1/2 from a for each additional step. The significance test to be performed are lower-tail tests, i.e., we reject H_o if the value of the observed ratio is smaller than the tabulated one. Tables of the lower percentage points of Beta are available in the Biometrika Tables for Statisticians, vol. 1 (Table 16). Notice there an exchange and doubling of parameters, i.e., $\nu_1 = 2b = p$ and $\nu_2 = 2a = (n_e-p+1)$, (n_e-p), (n_e-p-1), etc. This exchange was made in the Biometrika Tables to make parameters conform with the degrees of freedom in the associated F test.

Since the determinant of a matrix equals the product of leading coefficients, the likelihood-ratio and step-down tests are performed on the basis of the same numerical procedure, a forward Doolittle solution of the matrix (no right-hand side).

4.2 All Effects (Subtotals)

The matrix H + E is obtained by adding each element in the subtotal row of Table 5 to the corresponding one in the error row. We place the elements in the order 2-1-3, i.e.,

$$\begin{bmatrix} Y_2^2 & Y_1Y_2 & Y_2Y_3 \\ Y_1Y_2 & Y_1^2 & Y_1Y_3 \\ Y_2Y_3 & Y_1Y_3 & Y_3^2 \end{bmatrix}$$

For this and only this test (all effects) the elements of (H + E) are identical with those in the "Total" row of Table 5. Then

$$(H + E) \begin{bmatrix} 16{,}943.47 & 2{,}614.85 & 706.61 \\ 2{,}614.85 & 3{,}021.65 & 443.51 \\ 706.61 & 443.51 & 325.29 \end{bmatrix}$$

Applying forward Doolittle to $10^{-4}(H + E)$ we have

$$\begin{array}{rrr} 1.69435 & .26149 & .07066 \\ 1 & .15433 & .04170 \\ \hline & .30217 & .04435 \\ & .26181 & .03345 \\ & 1 & .12776 \\ \cline{2-3} & & .03253 \\ & & .02531 \end{array}$$

lead. coeff. of (H + E)	16,943.5	2,618.1	253.1
lead. coeff. of E (from Table 10)	323.9	30.7	184.8
β-ratio	.0191	.01172	.7301
Parameters (a, b)	(8, 3/2)	(15/2,3/2)	(7,3/2)
.01 values (Table 16 of Biometr. Tables)	.50194 very signif.	.47999 very signif.	.45615 non-signif.

Notice that we work on lower tail, i.e., β (observed) $< \beta$ (tabulated) indicates rejection of the null hypothesis.

The decision of the step-down analysis is:

Contribution of variable 2 very significant. Contribution of variable 1, over and beyond that explained by variable 2, still very significant. Contribution of variable 3, over and beyond that explained by 1 and 2, not significant.

For the likelihood-ratio test we compute the $\log_{10}$ of each leading coefficient in E and (H + E).

$$\log_{10} |E| = \log_{10} 323.9 + \log_{10} 30.7 + \log_{10} 184.8$$

$$= 2.51041 + 1.48714 + 2.26670 = 6.26425.$$

$$\log_{10} |H + E| = \log_{10} 16{,}943.5 + \log_{10} 2{,}618.1$$

$$+ \log_{10} 253.1 = 10.05028.$$

$$\log_{10} \Lambda = 6.62645 - 10.05028 = -3.42383.$$

$$m = 18 + \frac{1}{2} (5-3-1) = 18.5$$

$$n_h = 5 \quad f = 15$$

$$\delta = 15(9 + 25 - 5)/48 = 145/16.$$

$$-m \log_e \Lambda = (-2.302{,}585) \times 18.5 \times (-3.42383)$$

$$= 145.85$$

Table 7 of the Biometrika Tables lists $P < .00001$ for 15 d.f. and a value as large as 145.85. (It does not, in fact, go down, that far in degrees of freedom). The correction term involving δ/m^2 is even smaller, so that we need not go on. The null hypothesis of no over-all effect differences, for all variables combined, is very clearly rejected, as expected from the foregoing studies.

4.3 Interactions

In order 2-1-3,

$$H + E = \left[\begin{array}{ccc} 331.11 & 100.19 & 158.62 \\ 100.19 & 93.17 & 36.54 \\ 158.62 & 36.54 & 265.08 \end{array}\right]$$

(1280.06)

is obtained by adding elements in the (A x B) row to those of the E row in Table 5. The number in parentheses is obtained by the same method from the check column in Table 5, and must be equal to the sum of all elements in (H + E), except for rounding.

Forward Doolittle on 1/100 (H + E)

3.3111	1.0019	1.5862
1	.30259	.47906
	.9317	.3654
	.62854	- .11457
	1	- .18228
		2.6508
		1.87003

Lead. coeff. (H + E)	331.110	62.854	187.003
Lead. coeff. (E)	323.9	30.7	184.8
β	.9782	.4884	.9882
parameters	(8, 3/2)	(15/2, 3/2)	(7, 3/2)
.05 tabulated	.62217	.60333	.58256
.01 tabulated		.47999	

Here the values on variables 2 (first in set) and 3 (last in set) show that these variables are clearly non-significant. However, the contribution of variable 1 (middle term) is almost significant at the .01 level. There may thus be a slight contribution to interaction effect due to variable 1, which could be made stronger still by placing variable 1 first in this hierarchy. The significance is not nearly of the same strong character as that of all other effects studied thus far. Let us perform the over-all likelihood-ratio test:

$$\log_{10} |H + E| = \log_{10} 331.110 + \log_{10} 62.854$$
$$+ \log_{10} 187.003 = 6.59015$$

$$\log_{10} \Lambda = 6.626425 - 6.59015 = -.3259$$

$$m = 18 + \tfrac{1}{2}(2-3-1) = 17$$

$$f = 2 \times 3 = 6$$

$$\delta = 6\,(9 + 4 - 5)/48 = 1$$

$$-m \log_e \Lambda = -2.302,585\ m \log_{10} \Lambda = 12.757$$

$$\Pr[\chi^2_6 > 12.757] = .04725 \text{ (linear interpol.)}$$

$$\Pr[\chi^2_{10} > 12.757] = \underline{.23792} \text{('' '')}$$

$$\text{difference} \quad .19067$$

$$\delta/m^2 = 1/289, \text{ hence correction} = .19067/289 = .00066.$$

$$\text{Probability} = .00066 + .04725 = .04791\ .$$

The over-all test of interaction effects for all variables would thus reveal significance at just the .05 level. By contrast, the probability of the union-intersection test was about half-way between .05 and .01. This is the kind of difference which one may find between the two tests. If it is just around a pre-assigned significance level, different decisions may result. Significance level .05, if just barely attained, is of course no strong indication of a real effect difference. To obtain a closer look at this, we may look at the estimates of interaction effect of variable 1, which is clearly responsible for this situation.

$$\hat{\delta}_{11}^{(1)} = Y_{11.}^{(1)}/4 - \hat{\alpha}_1^{(1)} - \hat{\beta}_1^{(1)} - \hat{\mu}^{(1)}$$

$$= 40.375 + 3.3292 + 9.8209 - 54.8292$$

$$= -1.3041$$

The other estimates are obtained in the same manner, which is familiar from univariate analysis. This yields the following interaction effect estimates on variable 1:

$$\begin{array}{rr} -1.3041 & 1.3042 \\ 1.3709 & -1.3708 \\ -0.0666 & 0.0667 \end{array}$$

Contrasts of these interactions are of about the same order of magnitude as those of main effects in variable 3, which so far has not attained a single significant contribution. We would probably be led to the conclusion that there may be a slight interaction effect in variable 1, but certainly not enough to require abandonment of the two-way classification model.

4.4 A-Effect

In order 2-1-3, adding elements in row A and row E of Table 5, we have

$$H + E = \begin{bmatrix} 14{,}392.17 & 192.59 & 338.07 \\ 192.59 & 678.21 & 107.17 \\ 338.07 & 107.17 & 267.32 \end{bmatrix}$$

(16,613.37)

(for checks, see previous section).

Forward Doolittle on 1/1000 (H + E)

14.39217	.19259	.33807
1	.013882	.023490
	.67821	.10717
	.67563	.10265
	1	.15193
		.26732
		.24378

lead. coeff. (H + E)	14,392.17	675.63	243.78
" " E	323.9	30.7	184.8
β	.0225	.0454	.7580
$\beta_{.01}$	.50194	.47999	.45615
	very signif.	very signif.	not signif.

Conclusion: Same as that on all effects (Section 4.2)

m, f, δ as in interaction test (Section 4.3)

$\log_{10} \wedge = -3.11058$

$-m \log_e \wedge = 121.8$

compare with χ^2_6

$Pr < .00001$

very significant.

4.5 B-Effect

A likelihood-ratio test for the over-all multivariate hypothesis is not required, since for $n_h = 1$, the union-intersection and the likelihood-ratio tests reduce to the same statistic (Hotelling's T^2 or F-statistic).

For the step-down analysis, we may proceed in the customary manner. Adding, in variable order 2-1-3, the elements of row B and row E in Table 5 we obtain

$$(H + E) = \begin{bmatrix} 2,868.01 & 2,531.39 & 521.08 \\ 2,531.39 & 2,379.27 & 397.48 \\ 521.08 & 397.48 & 312.13 \end{bmatrix}$$

(12,495.32)

Forward Doolittle on 1/1000 (H + E)

$$\begin{array}{rrr} 2.86801 & 2.53139 & .52108 \\ 1 & .88263 & .18169 \\ \hline & 2.37927 & .39748 \\ & .14499 & -.06244 \\ & 1 & -.43065 \\ \hline & & .31213 \\ & & .19057 \end{array}$$

lead. coeff. (H + E)	2,868.01	144.99	190.57
" " E	323.9	30.7	184.8
β	.1129	.2117	.9697
$\beta_{.01}$	.50194	.47999	.45615
	very signif.	very signif.	non-signif.

Conclusion: Same as that on all effects (Section 4.2).

Multivariate Analysis of a Standard Two-Way Classification

Index

Analysis of covariance 1

Beta statistic

- step-down 24
- exchange of parameters 24
- table reference 24
- parameters, general 24
- all effects 25,26
- interaction 27
- A-effect 30
- B-effect 31

Centroid 18,19

Characteristic root, largest

- evaluation, reducing order 11,12,13
- direct evaluation 13
- distribution, standard form 13
  (see also Heck's charts)
- evaluation by iteration 18-22
- all effects 22
- interaction 13
- A-effect 14

Checking operations 3,4,5,6,7,8,12,19,27,29,31

Correction terms 6

Correlations 9

Covariances 8

Degrees of freedom 7,8

Determinants, evaluation of 25

Discriminant functions

- all effects 17,22
- A-effect 14,15
- B-effect 16

Distance, Mahalanobis 16,17

$E = \tilde{T}\tilde{T}'$, evaluation of $\tilde{T}^{-1}$ 17

Eigenvector, evaluation by iteration 18-22

Eigenvector(cont.)
    unspecified length 15
    initializing 18
    iteration 19-20
    Gauss-Seydel method 21
    comments on squaring a matrix 19

E-Matrix
    order 1-2-3 8
    order 2-1-3 10
    inversion 10
    leading coefficients 10,25,27,30, 31.

Estimation
    main effects 4,5
    interactions 29
    variances 8
    covariances 8
    correlations 9

Examples for response variables 1

Experimental units 1

F-test in union-intersection 16

Grand totals 5

$H = TT'$, method of obtaining T' 12
    T', for interaction 12
    T' for A-effect 14

Heck's charts
    reference 11
    parameters 13
    interaction 13
    A-effects 14
    all effects 22

H-matrix
    interaction 11
    rank of 11
    A-effects 14
    all effects 18

H + E matrix
    all effects 25
    interaction 27
    A-effect 29
    B-effect 31

Hotelling's T^2 30

Inversion 10

Leading Coefficients

  definition 10
  of E 10,25,27,30,31.
  square-roots of E 10
  of H + E, general 24
  of H + E, all effects 25
  of H + E, interaction 27
  of H + E, A-effect 30
  of H + E, B-effect 31

Likelihood-ratio tests

  general 23
  parameters 23
  detailed computation 26
  all effects 26
  interaction 27,28
  A-effect 29,30
  B-effect, reasons for not given 30

n_e 8

n_h 11

Observation vectors 2

Ordering of variables 9

Order (2-1-3),reasons 9

Principal components

  definition 18
  evaluation 18-21
  (see also Eigenvector)

Step-down technique

  general 24
  all effects 25,26
  interaction 27,28
  A-effect 30
  B-effect 31

Sum of squares and products

  totals, uncorrected 2
  A-effect 6,7
  B-effect 7
  Subtotals 6,7

Index

Sum of squares and products (cont.)
    Totals, corrected 6,7
    Interaction 6,7
    Error 6,7
    Signs of 6

Subtotals 2

Triangular matrices 10

    evaluation 10,17

Typical data for two-way 1

Union-intersection tests

    general 11
    interaction 11-13
    A-effect 13-15
    B-effect 16-17
    all effects 17-22

Wilks' lambda 23

APPENDIX B

A COMPARISON OF THE MAXIMUM DETERMINANT SOLUTION IN FACTOR ANALYSIS WITH VARIOUS APPROXIMATIONS

by

Robert H. Brown

The Market Review of

PEEP AND MOO

Virginia Polytechnic Institute and the United States Department of Agriculture Cooperating:
Extension Service, L. B. Dietrick, Director, Blacksburg, Virginia
Issued and Distributed in Furtherance of the Acts of Congress of May 8 and June 30, 1914

INTRODUCTION

Soon after the formulation of the "Multiple Factor" Model by L. L. Thurstone and Sir Godfrey Thomson, D. N. Lawley [5][1] presented a maximum-likelihood method for the estimation of factor loadings. His solution was in the form of an implicit equation which was impracticable to solve with the equipment then available. Bartlett [3] recommended a test which, however, assumed that the maximum-likelihood solution was known. Various approximations have been used with success, notably Thurstone's Centroid Method [7] and Thomson's approximation [6], but there are extremely few studies in the literature which report the maximum-likelihood solution. For this reason, the quality of approximation of the various methods is virtually unknown.

On intuitive grounds, L. L. Thurstone, in 1953, suggested a method of estimation which would maximize the ensuing partial correlation determinant. This method has been applied to several studies at the Psychometric Laboratory in Chapel Hill, North Carolina and in Frankfurt, Germany (see, e. g. [2]). Again, due to the lack of high speed computing equipment, no final solution to this "maximum-determinant" method was given. Howe [4] reported on this method and showed that it yields estimates identical to those obtained in the maximum-likelihood iteration, but neither are his examples carried to a final solution. Bargmann [1] uses the partial correlation determinant as a goodness-of-fit statistic and, in his "classification of dependence patterns", suggests a method which is analogous to Fisher's F-test used in curve-fitting by polynomials.

---

[1] Numbers in square brackets refer to the Bibiography.

In this thesis, we used the iterative technique for determination of maximum-likelihood (maximum-determinant) estimates as suggested in [1,4] and described in section 1.1. The factor matrix which one obtains in this iterative process is, however, not unique. As formula (1.1.3) shows, for any F satisfying the equation, there exist infinitely many equivalent solutions obtainable by multiplying F by any orthogonal matrix. To ascertain whether the iterative technique is leading to convergence (which, in theory, it always will [1]), we could select a particular solution at every stage. For some comparisons this was actually done, and we chose Thurstone's Principal Axes solution [7]. Of course, we were free to take any particular solution we pleased. After all, the least squares solution of a polynomial may be reported as a power series, a series in orthogonal polynomials, a series in factorial powers, or any other representation one may wish to choose, without changing the result. The same is true here (see, e.g. [1]). We did not use the very neat "Simple Structure" solution because it may not be uniquely determinate nor, for that matter, exist in every case. Lawley's Model I solution [5], which is also reported, is not particularly suitable in this instance (see Chapter II).

However, it is not necessary for us to select a particular solution at every stage of the iterative process. The equivalence of the maximum-likelihood and the maximum-determinant solutions permit us to take the determinant of the resulting partial correlation matrix as a criterion of convergence. If, after a number of iterations, we obtain a largest value for this determinant, i.e., if on subsequent iterations its value oscillates within narrow limits, we know we have a "maximum-determinant" and hence, a maximum-likelihood solution. So long as the value of the determinant increases from one iteration to the next, as it invariably does in the earlier stages, we cannot attribute the discrepancies to rounding errors, even if the increments are very small. The series of computer programs which we used in the present demonstration studies operate in six place, fixed-decimal numbers, so that rounding errors could be expected in the fifth, or possibly the fourth, decimal place. In the solutions obtained, the discrepancies in the value of the determinant were, indeed, of that order of magnitude, when oscillation occurred. Frequently, at much earlier stages, the increments were no greater than 10^{-5}, yet they were all positive. The first study will show what would happen if one prematurely adopted an early iteration as a final solution.

After we developed the rather intricate computer programs, we applied them to several actual studies; this thesis reports on three such studies. Study I, which was supplied by T. E. Jeffrey of the Psychometric Laboratory in Chapel Hill, contains 24 variables (psychological tests) and was based on a sample of N = 211. Study II, which was supplied by J. Ross Macdonald of Texas Instruments, Inc., contains 25 anthropological variables (body measurements) and was based on a sample of N = 350. Study III was obtained from the same source, has 12 variables, and is based on a sample size of 138.

To determine the number of factors in each case, we used a test originally proposed by Bartlett [3] and justified as a "goodness-of-fit" test by Bargmann [1]. The preliminary decision regarding the appropriate number of factors was made at the earliest possible stage (second centroid solution in Study I, first iteration of (1.1.4) in Study II, and fourth iteration of (1.1.4) in Study III). In no instance in the studies here reported, or in others which we undertook, did the final decision on the number of factors differ from the preliminary one. To illustrate, we must know the value of the determinant of the partial correlation matrix, given an estimate of the factor matrix F, in order to ascertain, by Bartlett's X^2-test, whether the k "artificial variables" or "factors" are sufficient to explain all dependence in the original correlation matrix [1]. However, as the iterative process continues, this determinant increases, the artificial variables produce a closer fit, and hence it is theoretically possible that, in the final solution, fewer than the k factors called for by the preliminary decision will be found to be sufficient. Similarly, we may think of using Fisher's F-test in trend-fitting if, for instance, we used the variate-difference method to determine the degree of the trend line. As we proceed from the variate-difference solution to the least squares solution, the fit will be closer in the sense of Fisher's goodness-of-fit test, and the degree may be less than that determined in the preliminary decision. No such change of decision was ever required in these factor analysis studies.

Along with the stabilizing of the partial correlation determinant, we expect stabilization of the communalities and of the factor loadings of the particular solution when the iterative process reaches convergence. But, of course, the order of magnitude of rounding errors is not the same for these results, so we decided, in line with our original intention, to conclude the process when the determinant indicated convergence had been reached.

In view of the slowness of the iterative process, a good estimate of the factor loadings is essential for an initial solution with which to enter (1.1.4). We used the centroid, or a once refactored centroid [7], as an initial estimate, provided it was "admissible" [1]. This, among other things, requires that each communality be less than unity. When this requirement was not met, we made slight reductions in the factor loadings of the appropriate row or rows so that we might be able to start the iterative process. In these instances (Studies II and III), we did not perform the preliminary test until the iterative procedure had produced an admissible solution for F, i.e., a positive-definite $I-F'R^{-1}F$ matrix. Prior to making the test, we obtained as many factors as the computer program for the centroid method would deliver.

This complete procedure is described in Chapter I and a discussion of the numerical results is given in Chapter II.

The following programs were written for the IBM 650 electronic computer and used in making the calculations for these demonstration studies.

Designation	Purpose
VPI 010	Obtain complete R matrix from input for initial centroid program, AM 557.
G 864 (5.2.001.2)	Obtain R^{-1} and $\|R\|$ in floating decimal form (North Carolina State College).
VPI 002	Obtain fixed decimal form of R^{-1}.
AM 557	Obtain initial centroid estimate of F matrix (Texas A. and M. College).
VPI 011	Replace principal diagonal elements of R with the communalities of the initial factor matrix approximation.
AM 561	Obtain improved centroid factor matrix.
VPI 012	Obtain final F matrix, by an iterative technique (maximum-determinant solution).
VPI 013	Obtain $I-F'R^{-1}F$ matrix.
VPI 014	Obtain $\log \|I-F'R^{-1}F\|$.
VPI 015	Obtain $\Sigma_{1} \log(1-h_i^2)$.
VPI 016 (w/VPI 012)	Obtain $F'F$ and/or $F'D_{1/(1-h^2)}F$ matrices.
5.2.005.2	Obtain characteristic roots (D_λ) and eigenvectors (E) of the matrix (matrices) punched by VPI 016 (by W. W. Piper).
VPI 017	Punch D_λ and E from 5.2.005.2.
VPI 018	Obtain FE matrices (P and L).

These programs have now been condensed and given library classification numbers. For a detailed description, see Rolf E. Bargmann and Robert H. Brown, "IBM 650 Programs for Factor Analysis", Virginia Polytechnic Institute, July 1961.

CHAPTER I

DISCUSSION OF PROCEDURE

1.1 Computational scheme.

The maximum-determinant approach and its computing formulae have been fully derived previously [1,4] but some of the conclusions given by Bargmann [1] are reproduced below for convenience.

Given a correlation matrix exhibiting the dependence pattern among p variates, it is desired to find a new set of k (artificial) variates which account for the observed correlations. To achieve this aim, it will be necessary to find a matrix F ($p \times k$; $k < p-1$), such that the correlation matrix

$$(1.1.1) \qquad D_{1/\sqrt{1-\underline{f}_i'\underline{f}_i}} \, (R - FF') \, D_{1/\sqrt{1-\underline{f}_i'\underline{f}_i}}$$

(where R is the given correlation matrix and $\underline{f}_i'$ is the i'th row vector of F)

approaches the identity matrix. In fact, F should be chosen in such a way that the determinant of (1.1.1) is maximized. Thus we proceed to maximize

$$(1.1.2) \qquad \phi(F) = \prod_{j=1}^{p} \frac{1}{1-h_j^2} |R - FF'|$$

(where $h_j^2 = \underline{f}_j'\underline{f}_j$)

under variation of F. When F is chosen in this manner, then the iterative formula

$$(1.1.3) \qquad (R - FF') \, D_{1/(1-h_i^2)} F = F$$

follows, provided $R-FF' \neq 0$ and all h_i^2 are less than unity. A modification of this last result may be made, to yield the more practical computing formula

$$(1.1.4)\quad \sum_{j\neq i=1}^{p} r_{ij}\frac{f_{j\alpha}}{1-h_j^2} = \sum_{\beta=1}^{p} f_{i\beta} \sum_{j\neq i=1}^{p} \frac{f_{j\alpha}}{1-h_j^2} f_{j\beta}\,,$$

$$i = 1,2,\ldots,p$$

$$j = 1,2,\ldots,k$$

The F matrix satisfying the equations (1.1.4) has been called the "maximum Δ" estimate. Proof has been given in the same paper [1] that the matrix obtained by iterating (1.1.4) will maximize Δand not simply produce a stationary value. Proof has also been given in [4] that the maximum-likelihood solution for the classical factor analysis model is equivalent to this "maximum Δ" solution.

1.2 Estimation of the number of factors.

The first step in the procedure is to apply the centroid method to obtain an initial guess of a factor matrix,F. In the computer program used for this purpose, the communalities (h_i^2) are taken to be the largest absolute values in the columns of the correlation matrix, as is customary. Factorization is continued until the sums of squares of the f_{ij} are equal to the sum of the initial communality estimates. The communalities for the F matrix so produced are then calculated and used in obtaining an improved factor matrix. That is, the centroid method is again applied, with the input correlation matrix now having principal diagonal elements equal to the communalities of the initial F matrix, since these serve as better estimates than those arbitrarily selected in the previous operation. This process is known as "refactoring", see [7].

At this point, "an attempt is made" to obtain log $|I-F'R^{-1}F|$ and $\Sigma\log(1-h^2)$ in order to perform a test on the adequacy of the number of factors. It is not always possible to obtain these values, as we shall see in 2.2. But, if they can be got, a test for "reduction to independence" can be performed as described in [1,5,7]. We test

(1.2.1) $\chi^2_\alpha(\nu) = -m_k \log_e |R(\underline{z}|\underline{x})|$

where the degrees of freedom are found by

(1.2.2) $\nu = \frac{p(p-1)}{2}$,

and the multiplier is given by Bartlett [3]

(1.2.3) $m_k = N - k - \frac{2p + 11}{6}$

(where N is the number of observations).

(The other terms in the χ^2 series given by Bartlett are quite small for the values of N customarily involved.) The determinant is obtained from (see [1])

(1.2.4) $\log_{10} |R(\underline{z}|\underline{x})| = \log_{10}|R| - \log_{10}|I-F'R^{-1}F|$
$-\Sigma\log_{10}(1-h^2)$,

and $R(\underline{z}|\underline{x})$ stands for the matrix of partial correlations between the observed variates ($\underline{z}$) given the artificially constructed ones ($\underline{x}$).

In accordance with suggestions made in [1], we determine the probability associated with each computed χ^2 value, for k = 1, 2,... factors. This sequence of probabilities had a very sharp "turning point" in all studies previously and presently investigated. We choose that value of k which gives rise to this "turning point" as being the appropriate number of factors (see Table 3.5.4).

This decision is identical with decisions usually made in goodness of fit tests, as shown in [1]. It must be kept in mind, however, that this decision is a preliminary one since we do not know, at this stage, the maximum-determinant solution. After obtaining it, we may, theoretically, have to modify the decision and restart the iterative process, which depends on the number of factors selected. However, in all studies thus far investigated, no such change of decision has ever been required.

1.3 The maximum-determinant solution.

After determining by the χ^2-test described in the previous section, that k factors are, with a given probability, sufficient to explain all the correlations, we apply (1.1.4). The improved factor matrix obtained by the centroid method is used as a starting estimate for the iterative procedure. The Gauss-Seidel method is applied in that the newly calculated $\underline{f}'_i$ is substituted for its predecessor before proceeding to the next row.

At intervals which seem appropriate, a complete F matrix is obtained and we perform the calculations necessary to make the convergence test. When convergence is indicated by "oscillation" of the value of log $|I-F'R^{-1}F| - \Sigma\log(1-h^2)$, as discussed in the Introduction and in section 2.3, an F matrix is chosen for which $|R(\underline{z}|\underline{x})|$ is a maximum. The χ^2-test is again applied to ascertain whether the preliminary decision was correct. The final factor matrix may then be rotated into the Principal Axes form of solution, or some other solution suitable to the particular problem.

1.4 Unique Solutions: Lawley's "Model I" and Thurstone's "Principal Axes".

It has been shown in [1,4] that iterating (1.1.4) to convergence does not yield a unique factor matrix F. Hence it is necessary, for purposes of making comparisons, to obtain the solutions in a common form. Lawley's Model I might be chosen to provide such a common basis. In this method, a unique solution is determined by diagonalizing the matrix $F'D_{1/(1-h^2)}F$. To do this, the eigenvectors are obtained, for then we have

(1.4.1) $E'F'D_{1/(1-h^2)}FE = D_\lambda$

(where E is the matrix of eiganvectors),

say. Since E is the eigenvector matrix, EE' = I and so, when postmultiplied by E, (1.1.3) may be written as

(1.4.2) $(R - FEE'F')\ D_{1/(1-h^2)}FE = FE.$

Letting FE = L, Lawley's form of the solution, we have

(1.4.2.a) $(R - LL')\ D_{1/(1-h^2)}\ L = L.$

It follows that the communalities are invariant under this transformation, for

(1.4.3) $h_i^2 = (FF')_{ii}$

and

(1.4.4) $LL' = FEE'F' = FF'.$

Thus L is itself a solution to the iterative equation.

Another form of solution, which we chose as a common standard for making comparisons, is Thurstone's "Derived Principal Axes". (This should not be confused with the Principal Components of R, a quite different kind of matrix). It merely is that solution, P, of the equation

(1.4.5) $(R - PP')\ D_{1/(1-h^2)}\ P = P$

for which P'P is diagonal. It is obtained in the same way as L, except that we take the eigenvectors of F'F instead of $F'D_{1/(1-h^2)}F$. This "Principal Axes" solution is unique and provides a suitable basis for determination of the adequacy of our approximations.

CHAPTER II

COMPARISON OF ALTERNATIVE SOLUTIONS

2.1 Study I (FA 1): differences among the centroid result and the fifth and forty-second iterations.

Data tables for this study (and the other studies) may be found in the Appendices. The coding system for the tables was formulated as follows: the first digit is the study number, the second digit represents the type of information contained in the table and the third digit is the page number for the particular set of data.

TABLES: 1.1.1-3, Correlation matrix R; 1.2.1-5, F matrices; 1.3.1-4, Communalities; 1.4.1-4, $I-F'R^{-1}F$ matrices; 1.5.1, Convergence test; 1.5.2, Determination of number of factors; 1.6.1, Characteristic roots of $F'F$; 1.7.1-3, $FE_1=P$ matrices; 1.8.1-3, $FE=L$ matrices; 1.9.1, Frequency list of differences among P matrices.

Table 1.5.1 gives evidence of the very slow rate of convergence mentioned in the Introduction. Between the 4th and 6th iterations, the determinant was increasing by very small amounts (less than 0.0001) but it was still monotonically increasing, so the discrepancies could not be attributed to rounding errors (which are expected in the fifth decimal place). It took 42 iterations to arrive at the maximum-determinant solution. The attainment of the maximum was signalled by fluctuation of the discrepancies, which could then be interpreted as rounding errors. One might have been tempted to halt after five iterations but comparison of the results shows that the fifth iteration is no better than the centroid as an approximation.

A tabulation of the differences among the P matrices (Thurstone's Principal Axes solutions) has been made and a summary of the results is presented in Table 1.9.1. It can be seen that after five iterations, the loadings[2] on the first factor are still almost identical to those of the centroid. While there are numerous small differences on the other factors, none of these have, of course, loadings comparable in size to those of the first factor. In fact, the fifth iteration, which showed "pseudo-convergence", is actually much closer to the centroid result than to the maximum-determinant solution.

The differences between the centroid and the final, maximum-determinant solution are also quite small. For the first factor, only one difference reaches 0.02 and most discrepancies were 0.01 or less. The spread is somewhat greater for the remaining factors but, even so, differences of 0.10 are rare and 0.111 was the largest observed.

Lawley's form of the solution would have been a desirable basis for comparison, but the communalities are so close to unity in some instances that the results are distorted. This is due to the fact that if a communality increases from, say, 0.90 to 0.99, the divisor $1-h^2$ decreases by a factor of 10. Of course, if Simple Structure is desired, F, P, and L will yield the same structure, for a given solution, since they are linear transforms of each other. The Lawley results (L matrices, Tables 1.8.1-3) are also reported for completeness.

---

[2] The word "loadings" simply denotes the elements of the factor matrix, here P.

2.2 Study II (FA III): Differences between the centroid result and the twelfth iteration.

TABLES: 2.1.1-3, Correlation matrix R; 2.2.1-8, F matrices; 2.3.1-2, Communalities; 2.4.1-4, $I-F'R^{-1}F$ matrices; 2.5.1, $\log |I-F'R^{-1}F|$ and $\Sigma \log(1-h^2)$; 2.5.2, Convergence test; 2.5.3, Determination of the number of factors; 2.6.1, Characteristic roots of $F'F$; 2.7.1-4, $FE_1=P$ matrices; 2.8.1-4, $FE=L$ matrices; and 2.9.1, Frequency list of differences between P matrices.

The variations between the Principal Axes solutions for this problem are seen to be slight. For the first factor, only one difference was as great as 0.02. For the other factors, the discrepancies were somewhat larger, but the maximum was 0.174. The characteristic roots indicate the hierarchy of the factors and, in this case, there seems to be an interchange of factors. For the twelfth iteration, the 6th, 7th and 8th roots are very close. The P matrices suggest that factor VII of the 12th iteration is factor VI of the centroid, with opposite signs. In the summary of differences between solutions, this interchange of factors was made before computing the variations.

As in Study I, the F matrices appear quite similar, except for a few variables such as 8, 22, and 23. These same variables show sharp trends in the communalities. Another item worthy of note here is that the centroid operation produced one communality, that for variable 11, which was greater than unity. Refactoring by the centroid method reduced this communality but still failed to produce an admissible factor matrix [1]. Therefore, it was necessary to scale down that row of F before attempting to use the centroid result as the input for the iterative procedure. As is pointed out in the reference above, altering the communality estimate does not affect the final value of the communality when convergence is reached.

It might be further noted that the existence of a communality greater than unity made it impossible to obtain the $I-F'R^{-1}F$ matrix and, of course, $\log |I-F'R^{-1}F|$. Hence these items are missing from Tables 2.4.1 and 2.5.1-2. The modified centroid matrix was used to obtain the particular solutions.

Just as in Study I, some communalities show sharp trends, either approaching or receding from 1.00, so that, again, the Lawley form of solution is unsuitable as a basis for making comparisons. It is included, however, for the sake of completeness.

2.3 Study III (FA 116): differences among the centroid result and the fortieth and fifty-ninth iterations.

TABLES: 3.1.1, Correlation matrix R; 3.2.1-2, F matrices; 3.3.1-2, Communalities; 3.4.1-3, $I-F'R^{-1}F$ matrices; 3.5.1, $\log |I-F'R^{-1}F|$; 3.5.2-3 $\Sigma \log(1-h^2)$;3.5.4 number of factors; 3.6.1, Characteristic roots of $F'F$; 3.7.1, $FE_1=P$ matrices; 3.8.1, $FE=L$ matrices; 3.9.1, Frequency list of differences among P matrices.

To illustrate how the determinant oscillates when convergence is reached, 64 iterations were performed in this problem. The 59th was chosen as the convergent iteration because it had a value for the determinant which was about in the center of the oscillations. Also, the smallest increment occurred between it and the 60th iteration. The 40th iteration was selected to be compared with the exact solution to provide a contrast with Study I, where an early iteration was picked.

As in the first two studies, the Principal Axes results show that the centroid method provides quite a good approximation to the maximum-determinant solution, (But the Lawley result is again unsuitable for making comparisons, for the same reason as before.) No variations greater than 0.01 were observed in the P matrices, for the first factor. Among the other factors, the greatest discrepancy recorded was 0.158 and only four differences exceeded 0.10.

The 40th iteration shows an even better correspondence with the exact solution. However, study of Table 3.5.3 reveals that there was no particular reason to pick this iteration over others, which might have provided equally good approximations. And the other; intermediate results: the communalities and the $I-F'R^{-1}F$ matrices, which one might hope would serve as indicators of satisfactory convergence, do not, on the evidence of these studies, serve the purpose. This problem does seem to indicate that the later iterations would provide highly satisfactory approximations to the maximum-determinant solution, _if_ one were able to discover when the process had progressed far enough. Even the convergence test of Table 3.5.3 does not appear to serve as an adequate indicator so that additional research will probably be required to provide an answer to this question.

2.4 Summary and Conclusions.

If the close approximation of the refactored centroid is not sufficient for the researcher's purposes (e.g., if he has very accurate measurements on a very large sample) he will not gain much by making just a few iterations of the maximum-determinant procedure. Our experience has shown that the earlier iterative solutions are closer to the centroid than to the final solution. However, where the communalities are very large, the centroid may produce an inadmissible F. In such cases, a few iterations may be sufficient to allow the decision to be made on the number of factors while also producing a slight improvement in the factor matrix approximation. And, since we now have computer programs to perform the maximum-determinant iterations, there is no reason to compromise if great accuracy is desired, provided the proper equipment is available.

BIBLIOGRAPHY

[1] Bargmann, R.E., "A Study of Independence and Dependence in Multivariate Normal Analysis", Institute of Statistics, University of North Carolina, No. 186 (1957).

[2] ________, "A Demonstration Study on the Effectiveness of Factor-Analytical Methods", Hochschule f. Internationale Paed. Forschung, Frankfurt/Main, Forschungsbericht (June 1955).

[3] Bartlett, M.S., "Tests of Significance in Factor Analysis", *British Journal of Psychology* (Statistical Section) vol. III (1950) page 77.

[4] Howe, W. G., *Some Contributions to Factor Analysis*, Oak Ridge National Laboratory, ORNL No. 1919 (1955).

[5] Lawley, D.N., "The Estimation of Factor Loadings by the Method of Maximum Likelihood", *Proc. of the Royal Society of Edinburgh* vol. LX (1940) page 64.

[6] Thomson, Sir Godfrey, *The Factorial Analysis of Human Ability*, Houghton Mifflin, Boston (1939).

[7] Thurstone, L.L., *Multiple Factor Analysis*, University of Chicago Press, (1940).

```
      FA 1     R    CORRELATION MATRIX -- 6 DECIMAL PLACES

ROW  COL
  1  01    1000000   392000   549000   352000   600000   560000   475000
  1  08     343000   546000   634000   541000   364000   349000   420000
  1  15     534000   496000   255000   331000   433000   428000   584000
  1  22     584000   349000   514000
  2  01     392000  1000000   252000   271000   338000   365000   255000
  2  08     158000   367000   338000   381000   264000   173000   230000
  2  15     328000   396000   409000   354000   210000   412000   436000
  2  22     397000   288000   276000
  3  01     549000   252000  1000000   231000   486000   472000   437000
  3  08     440000   454000   457000   448000   423000   468000   525000
  3  15     427000   331000   304000   324000   393000   322000   441000
  3  22     419000   364000   394000
  4  01     352000   271000   231000  1000000   276000   261000   244000
  4  08     245000   274000   414000   296000   231000   178000   312000
  4  15     245000   269000   405000   423000   242000   363000   295000
  4  22     454000   293000   283000
  5  01     600000   338000   486000   276000  1000000   627000   340000
  5  08     389000   611000   543000   557000   287000   348000   406000
  5  15     671000   461000   359000   405000   373000   423000   505000
  5  22     493000   342000   492000
  6  01     560000   365000   472000   261000   627000  1000000   387000
  6  08     291000   556000   554000   420000   389000   380000   370000
  6  15     594000   441000   338000   454000   528000   462000   427000
  6  22     496000   275000   448000
  7  01     475000   255000   437000   244000   340000   387000  1000000
  7  08     295000   404000   424000   302000   296000   256000   257000
  7  15     355000   277000   240000   268000   369000   285000   362000
  7  22     392000   229000   357000
  8  01     343000   158000   440000   245000   389000   291000   295000
  8  08    1000000   304000   316000   326000   359000   431000   321000
  8  15     215000   310000   222000   254000   213000   223000   338000
  8  22     259000   213000   314000

DETERMINANT R EQUALS   .000002834
```

TABLE 1.1.1

```
            FA 1       R    CORRELATION MATRIX -- 6 DECIMAL PLACES
ROW  COL
  9   01     546000    367000    454000    274000    611000    556000    404000
  9   08     304000   1000000    480000    538000    313000    355000    429000
  9   15     598000    447000    320000    471000    445000    435000    576000
  9   22     444000    259000    477000
 10   01     634000    338000    457000    414000    543000    554000    424000
 10   08     316000    480000   1000000    444000    303000    360000    339000
 10   15     446000    349000    302000    355000    466000    444000    451000
 10   22     688000    319000    370000
 11   01     541000    381000    448000    296000    557000    420000    302000
 11   08     326000    538000    444000   1000000    243000    345000    471000
 11   15     520000    307000    324000    372000    340000    420000    805000
 11   22     466000    245000    410000
 12   01     364000    264000    423000    231000    287000    389000    296000
 12   08     359000    313000    303000    243000   1000000    427000    414000
 12   15     249000    245000    197000    298000    396000    334000    292000
 12   22     292000    314000    346000
 13   01     349000    173000    468000    178000    348000    380000    256000
 13   08     431000    355000    360000    345000    427000   1000000    426000
 13   15     268000    332000    259000    284000    465000    352000    402000
 13   22     353000    327000    415000
 14   01     420000    230000    525000    312000    406000    370000    257000
 14   08     321000    429000    339000    471000    414000    426000   1000000
 14   15     371000    398000    371000    393000    389000    461000    468000
 14   22     445000    385000    385000
 15   01     534000    328000    427000    245000    671000    594000    355000
 15   08     215000    598000    446000    520000    249000    268000    371000
 15   15    1000000    445000    286000    387000    402000    453000    560000
 15   22     478000    332000    491000
 16   01     496000    396000    331000    269000    461000    441000    277000
 16   08     310000    447000    349000    307000    245000    332000    398000
 16   15     445000   1000000    423000    417000    379000    440000    432000
 16   22     442000    295000    486000
```

TABLE 1.1.2

```
FA 1       R    CORRELATION MATRIX -- 6 DECIMAL PLACES

ROW  COL
 17  01      255000    409000    304000    405000    359000    338000    240000
 17  08      222000    320000    302000    324000    197000    259000    371000
 17  15      286000    423000   1000000    410000    284000    446000    316000
 17  22      323000    359000    332000
 18  01      331000    354000    324000    423000    405000    454000    268000
 18  08      254000    471000    355000    372000    298000    284000    393000
 18  15      387000    417000    410000   1000000    310000    471000    383000
 18  22      380000    317000    437000
 19  01      433000    210000    393000    242000    373000    528000    369000
 19  08      213000    445000    466000    340000    396000    465000    389000
 19  15      402000    379000    284000    310000   1000000    494000    432000
 19  22      475000    349000    352000
 20  01      428000    412000    322000    363000    423000    462000    285000
 20  08      223000    435000    444000    420000    334000    352000    461000
 20  15      453000    440000    446000    471000    494000   1000000    506000
 20  22      539000    440000    428000
 21  01      584000    436000    441000    295000    505000    427000    362000
 21  08      338000    576000    451000    805000    292000    402000    468000
 21  15      560000    432000    316000    383000    432000    506000   1000000
 21  22      529000    390000    468000
 22  01      584000    397000    419000    454000    493000    496000    392000
 22  08      259000    444000    688000    466000    292000    353000    445000
 22  15      478000    442000    323000    380000    475000    539000    529000
 22  22     1000000    473000    438000
 23  01      349000    288000    364000    293000    342000    275000    229000
 23  08      213000    259000    319000    245000    314000    327000    385000
 23  15      332000    295000    359000    317000    349000    440000    390000
 23  22      473000   1000000    336000
 24  01      514000    276000    394000    283000    492000    448000    357000
 24  08      314000    477000    370000    410000    346000    415000    385000
 24  15      491000    486000    332000    437000    352000    428000    468000
 24  22      438000    336000   1000000
```

TABLE 1.1.3

TABLE 1.2.1

```
                FA 1      F MATRICES          6 DECIMAL PLACES

 F AFTER CENTROID              AFTER 5 ITERATIONS           AFTER 42 ITERATIONS

FACTR VAR   ELEMENT        FACTOR VAR   ELEMENT         FACTOR VAR   ELEMENT

  1   01     752479            1   01     748778             1   01     748193
  1   02     513761            1   02     508592             1   02     508071
  1   03     662427            1   03     655179             1   03     655093
  1   04     484221            1   04     474731             1   04     479793
  1   05     735593            1   05     734343             1   05     735268
  1   06     715014            1   06     715987             1   06     715695
  1   07     524368            1   07     516484             1   07     517478
  1   08     475733            1   08     468927             1   08     471289
  1   09     712058            1   09     712059             1   09     712104
  1   10     699629            1   10     704468             1   10     705947
  1   11     686956            1   11     691194             1   11     684680
  1   12     512694            1   12     505400             1   12     505377
  1   13     563903            1   13     563371             1   13     563433
  1   14     631337            1   14     634037             1   14     630799
  1   15     685230            1   15     693459             1   15     692758
  1   16     621829            1   16     616374             1   16     617293
  1   17     527531            1   17     515875             1   17     517187
  1   18     597337            1   18     590510             1   18     591687
  1   19     619806            1   19     619773             1   19     618958
  1   20     676597            1   20     679981             1   20     677801
  1   21     746298            1   21     747498             1   21     759214
  1   22     730572            1   22     732898             1   22     731131
  1   23     524002            1   23     527942             1   23     529633
  1   24     647328            1   24     645474             1   24     644771
```

TABLE 1.2.2

```
          FA 1       F MATRICES           6 DECIMAL PLACES

   F AFTER CENTROID              AFTER 5 ITERATIONS           AFTER 42 ITERATIONS

  FACTR VAR   ELEMENT         FACTOR VAR   ELEMENT          FACTOR VAR   ELEMENT

     2  01     235759-             2  01    226863-              2  01    296650
     2  02      91250              2  02    109017               2  02    114272
     2  03     153056-             2  03    126468-              2  03    117746-
     2  04     215558              2  04    222684               2  04    248328
     2  05     300578-             2  05    283024-              2  05    279655-
     2  06     126776-             2  06    137389-              2  06    138471-
     2  07     138001-             2  07     96185-              2  07    105059-
     2  08      56685-             2  08     55309-              2  08     49027-
     2  09     198126-             2  09    191999-              2  09    194430-
     2  10     147111-             2  10    150896-              2  10    145060-
     2  11     305809-             2  11    312217-              2  11    256308-
     2  12     131691              2  12    128956               2  12    124970
     2  13     100314              2  13     78000               2  13     74979
     2  14     140580              2  14    136263               2  14    152334
     2  15     244159-             2  15    242995-              2  15    256617-
     2  16     145700              2  16    139462               2  16    114677
     2  17     305629              2  17    303557               2  17    334846
     2  18     199019              2  18    203306               2  18    223818
     2  19      91395              2  19     74766               2  19     58410
     2  20     289999              2  20    278043               2  20    273194
     2  21     173592-             2  21    176496-              2  21    201632-
     2  22      74675              2  22     61052               2  22     55363
     2  23     268559              2  23    272801               2  23    248413
     2  24      25307              2  24     21921               2  24     12773
```

TABLE 1.2.3

```
                 FA 1       F MATRICES          6 DECIMAL PLACES

      F AFTER CENTROID                  AFTER 5 ITERATIONS              AFTER 42 ITERATIONS

    FACTR VAR    ELEMENT             FACTOR VAR    ELEMENT            FACTOR VAR    ELEMENT

        3  01      102369-                3  01      100369-                3  01      103455-
        3  02      241802                 3  02      185957                 3  02      185317
        3  03      297289-                3  03      262066-                3  03      260603-
        3  04       41992-                3  04       35193-                3  04       30384-
        3  05      154418                 3  05      149523                 3  05      155572
        3  06       66741                 3  06       42539                 3  06       41872
        3  07      164868-                3  07      153681-                3  07      152271-
        3  08      205336-                3  08      229508-                3  08      219951-
        3  09      172704                 3  09      168434                 3  09      169412
        3  10      232838-                3  10      272614-                3  10      268446-
        3  11      186712                 3  11      169285                 3  11      156795
        3  12      266106-                3  12      253780-                3  12      253817-
        3  13      244295-                3  13      273908-                3  13      275085-
        3  14       85124-                3  14       74002-                3  14       77398-
        3  15      261313                 3  15      291411                 3  15      269184
        3  16      190761                 3  16      159536                 3  16      161309
        3  17      176673                 3  17      159021                 3  17      166911
        3  18      203500                 3  18      180961                 3  18      186943
        3  19      148361-                3  19      128140-                3  19      133372-
        3  20      128100                 3  20      128418                 3  20      122414
        3  21      160848                 3  21      166496                 3  21      187329
        3  22      148431-                3  22      131394-                3  22      136076-
        3  23       81827-                3  23       51304-                3  23       45991-
        3  24       75338                 3  24       85161                 3  24       82881
```

TABLE 1.2.4

```
                      FA 1      F MATRICES          6 DECIMAL PLACES

  F AFTER CENTROID                 AFTER 5 ITERATIONS              AFTER 42 ITERATIONS

FACTR VAR    ELEMENT            FACTOR VAR    ELEMENT           FACTOR VAR    ELEMENT

  4  01         99789              4  01      108023               4  01      101326
  4  02         91516              4  02       61031               4  02       64785
  4  03       157218-              4  03     161697-               4  03     144062-
  4  04        175885              4  04      149233               4  04      185907
  4  05         50926              4  05      112576               4  05      136728
  4  06        219793              4  06      194878               4  06      175879
  4  07        113574              4  07       65957               4  07       44294
  4  08       216121-              4  08     210943-               4  08     204316-
  4  09         29135              4  09       2310-               4  09       7955-
  4  10        368970              4  10      369409               4  10      369667
  4  11       290072-              4  11     305376-               4  11     237656-
  4  12       153260-              4  12     162260-               4  12     171083-
  4  13       300575-              4  13     253160-               4  13     263193-
  4  14       236311-              4  14     236561-               4  14     192431-
  4  15         63880              4  15      103158               4  15       88062
  4  16         61527              4  16       57407               4  16       25819
  4  17         29555              4  17        6961               4  17       43767
  4  18         20606              4  18       24055               4  18       42477
  4  19         92929              4  19       33691               4  19        533-
  4  20         57291              4  20       53218               4  20       54260
  4  21       282439-              4  21     293764-               4  21     376333-
  4  22        243728              4  22      267750               4  22      271319
  4  23        31176-              4  23       9194-               4  23      31787-
  4  24        51499-              4  24      41883-               4  24      49683-
```

```
                 FA 1        F MATRICES           6 DECIMAL PLACES

     F AFTER CENTROID                AFTER 5 ITERATIONS              AFTER 42 ITERATIONS

   FACTR VAR   ELEMENT          FACTOR VAR    ELEMENT           FACTOR VAR     ELEMENT

     5   01      62834            5   01        47247             5   01          58028
     5   02     205836            5   02       144391             5   02         141746
     5   03     113127-           5   03       165870-            5   03         190938-
     5   04     199824            5   04       182838             5   04         136475
     5   05     104134-           5   05       159427-            5   05         194813-
     5   06     257061-           5   06       223555-            5   06         195654-
     5   07      40112-           5   07        46902-            5   07          17149-
     5   08     110827-           5   08       135609-            5   08         148428-
     5   09     151172-           5   09        96257-            5   09          87654-
     5   10     160599            5   10       176285             5   10         170755
     5   11     331359            5   11       341827             5   11         259265
     5   12     183707-           5   12       203750-            5   12         188919-
     5   13     176164-           5   13       142689-            5   13         132445-
     5   14      12719-           5   14        27623-            5   14          82055-
     5   15     100042-           5   15       109686-            5   15          88565-
     5   16     133832-           5   16       123018-            5   16          80002-
     5   17      37543            5   17         3216-            5   17          51879-
     5   18      78214-           5   18        66459-            5   18          90125-
     5   19     138240-           5   19        54814-            5   19           6283-
     5   20      75714            5   20        98817             5   20          99164
     5   21     321487            5   21       340783             5   21         442874
     5   22     260261            5   22       260758             5   22         255054
     5   23      91185            5   23        48248             5   23          75296
     5   24     147682-           5   24       132689-            5   24         120371-
```

TABLE 1.2.5

TABLE 1.3.1

```
                FA 1      COMMUNALITIES        6 DECIMAL PLACES

VAR    1          2          3          4          5          6          7    ITERATION

    643627     336046     561975     334708     657832     625045     313038   1
    638321     331664     564621     335794     673148     623013     307096   2
    637455     330345     567055     333379     677266     621683     306664   3
    636524     329831     567482     332301     678948     621306     306436   4
    636110     329704     567591     331896     679810     621277     306175   5
    635912     329698     567660     331769     680337     621325     305937   6
    635817     329731     567692     331770     680781     621374     305727   7
    635797     329774     567698     331842     681213     621410     305537   8
    635837     329829     567698     331960     681672     621408     305362   9
    635940     329886     567683     332109     682174     621353     305203  10
    636086     329945     567674     332287     682722     621236     305054  11
    636276     330002     567658     332502     683321     621040     304919  12
    636498     330062     567642     332758     683982     620740     304798  13
    637018     330170     567610     333411     685486     619806     304575  15
    637585     330277     567599     334267     687209     618381     304399  17
    638408     330380     567617     335836     690038     615494     304230  20
    639068     330405     567672     337513     692790     612251     304153  23
    639611     330316     567788     339399     695697     608465     304146  27
    639839     330204     567877     340406     697193     606340     304173  30
    639980     330086     567965     341109     698214     604793     304204  33
    640103     329892     568098     341929     699344     602914     304249  39
    640120     329876     568116     342021     699466     602708     304251  40
    640121     329850     568135     342088     699575     602526     304252  41
    640133     329826     568136     342156     699675     602361     304263  42
```

TABLE 1.3.2

FA 1 COMMUNALITIES 6 DECIMAL PLACES

VAR 8	9	10	11	12	13	14	ITERATION
325916	589679	746847	821775	401046	495059	487487	1
333188	584846	753710	822539	401854	486585	485143	2
336863	582504	756557	820495	403436	483848	483804	3
338120	581749	758949	817428	404121	483132	483163	4
338512	581532	760903	813986	404305	482947	482771	5
338661	581456	762447	810264	404345	482882	482400	6
338715	581435	763648	806258	404361	482873	482027	7
338716	581444	764564	801966	404355	482899	481614	8
338700	581471	765230	797395	404347	482945	481171	9
338663	581513	765672	792541	404328	483012	480683	10
338618	581564	765957	787405	404285	483099	480173	11
338554	581630	766075	782018	404224	483181	479622	12
338497	581689	766069	776405	404145	483289	479039	13
338353	581815	765737	764625	403906	483538	477816	15
338172	581911	765076	752519	403575	483819	476562	17
337867	581982	763665	735002	402954	484262	474813	20
337550	581937	762053	719832	402299	484695	473419	23
337182	581769	760029	704711	401564	485112	472201	27
336992	581642	758821	696902	401167	485303	471652	30
336851	581535	757900	691360	400877	485423	471313	33
336698	581394	756768	684712	400526	485536	470950	39
336691	581387	756661	683987	400482	485541	470919	40
336679	581363	756520	683345	400436	485539	470888	41
336671	581342	756431	682764	400406	485563	470866	42

TABLE 1.3.3

FA 1 COMMUNALITIES 6 DECIMAL PLACES

VAR 15	16	17	18	19	20	21	ITERATION
631599	441983	387380	432916	415172	569653	807520	1
642367	442896	385826	429420	411729	567914	811133	2
646092	443667	384908	428519	410726	568153	814003	3
647128	443564	383995	427991	410444	568550	816930	4
647525	443247	383620	427778	410268	568770	820055	5
647743	442852	383620	427826	410093	568866	823431	6
647852	442399	383865	428068	409905	568885	827099	7
647898	441883	384294	428450	409699	568865	831081	8
647882	441303	384869	428945	409473	568806	835413	9
647850	440674	385589	429547	409227	568722	840119	10
647775	439990	386428	430235	408960	568609	845214	11
647671	439272	387389	431003	408674	568460	850705	12
647560	438511	388476	431843	408366	568286	856605	13
647293	436923	390966	433677	407711	567840	869549	15
647002	435311	393795	435615	407044	567257	883717	17
646534	433046	398271	438412	406133	566205	905985	20
646096	431184	402386	440707	405454	565102	927221	23
645623	429466	406534	442766	404911	563811	950615	27
645383	428637	408617	443698	404686	563105	963779	30
645213	428084	410030	444291	404544	562594	973635	33
645040	427462	411634	444913	404388	561973	986114	39
645014	427398	411807	444977	404371	561922	987521	40
644999	427341	411950	445022	404363	561858	988791	41
644992	427289	412070	445062	404349	561812	989917	42

TABLE 1.3.4

```
              FA 1      COMMUNALITIES      6 DECIMAL PLACES

VAR  22       23       24                                        ITERATION
    701122   362131   444886                                        1
    700544   359605   443626                                        2
    699681   358750   443690                                        3
    698697   358417   443747                                        4
    697816   358188   443730                                        5
    697073   357976   443684                                        6
    696481   357746   443608                                        7
    696019   357503   443504                                        8
    695652   357250   443378                                        9
    695386   356971   443231                                       10
    695176   356682   443066                                       11
    695031   356379   442888                                       12
    694925   356065   442697                                       13
    694802   355410   442292                                       15
    694734   354729   441877                                       17
    694663   353749   441291                                       20
    694621   352899   440802                                       23
    694629   352075   440332                                       27
    694670   351674   440101                                       30
    694710   351404   439942                                       33
    694764   351097   439764                                       39
    694774   351069   439740                                       40
    694789   351044   439735                                       41
    694801   351015   439720                                       42
```

TABLE 1.4.1

```
       FA 1        I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
ROW  COL                    AFTER CENTROID SOLUTION
   1  01     38918      25220       5252-      1195-     16036-
   2  01     25221     310369      33413      34620-     27332
   3  01      5253-     33412     337692      33624      30237-
   4  01      1195-     34620-     33624     295034      36019
   5  01     16037-     27333      30237-     36019     319712
ROW  COL                    AFTER ITERATION 1
   1  01     42977      31524       2862-      3756-     24337-
   2  01     31525     341958      27903      25387-     38337
   3  01      2861-     27903     371586      45106      35049-
   4  01      3757-     25388-     45107     308287      38246
   5  01     24336-     38337      35049-     38246     341772
ROW  COL                    AFTER ITERATION 2
   1  01     42760      30642       4770-      2924-     21795-
   2  01     30643     342013      32698      25049-     33173
   3  01      4771-     32697     364708      46330      26995-
   4  01      2927-     25048-     46330     305437      40361
   5  01     21797-     33173      26995-     40360     334292
ROW  COL                    AFTER ITERATION 3
   1  01     42524      30856       4385-      2955-     21112-
   2  01     30856     340613      32908      24166-     32564
   3  01      4386-     32908     363384      46726      27713-
   4  01      2957-     24166-     46725     305216      40061
   5  01     21112-     32564      27713-     40061     332459
ROW  COL                    AFTER ITERATION 4
   1  01     42477      30919       4200-      2967-     21097-
   2  01     30919     340978      32860      23274-     32231
   3  01      4201-     32860     362774      47129      27497-
   4  01      2968-     23274-     47129     304616      39846
   5  01     21099-     32231      27497-     39847     332069
ROW  COL                    AFTER ITERATION 5
   1  01     42451      30955       4140-      2963-     21174-
   2  01     30956     341432      32641      22505-     31830
   3  01      4139-     32641     362218      47562      27135-
   4  01      2964-     22504-     47562     304101      39683
   5  01     21173-     31830      27135-     39683     331777
```

TABLE 1.4.8

```
          FA 1       I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
 ROW  COL                           AFTER ITERATION 6
   1  01          42424     30974      4117-      2955-     21277-
   2  01          30976    341911     32364      21802-     31435
   3  01           4116-    32364    361765      47970      26924-
   4  01           2955-    21801-    47970     303662      39620
   5  01          21275-    31435     26924-     39620     331459
 ROW  COL                           AFTER ITERATION 11
   1  01          42231     30922      4307-      2438-     22196-
   2  01          30923    344417     31254      19198-     29434
   3  01           4307-    31255    360462      49646      27307-
   4  01           2440-    19197-    49646     300765      41809
   5  01          22196-    29434     27307-     41808     327830
 ROW  COL                           AFTER ITERATION 12
   1  01          42175     30915      4406-      2219-     22489-
   2  01          30918    344899     31118      18907-     29155
   3  01           4405-    31118    360283      50020      27670-
   4  01           2221-    18907-    50020     299838      42954
   5  01          22489-    29154     27670-     42955     326411
 ROW  COL                           AFTER ITERATION 13
   1  01          42112     30918      4525-      1949-     22832-
   2  01          30920    345344     31006      18712-     28963
   3  01           4524-    31005    360099      50435      28156-
   4  01           1949-    18713-    50436     298685      44404
   5  01          22832-    28962     28156-     44404     324655
 ROW  COL                           AFTER ITERATION 15
   1  01          41965     30962      4824-      1257-     23678-
   2  01          30964    346097     30899      18677-     28919
   3  01           4825-    30898    359726      51444      29462-
   4  01           1256-    18677-    51445     295680      48333
   5  01          23677-    28919     29461-     48333     320013
 ROW  COL                           AFTER ITERATION 17
   1  01          41789     31069      5201-       357-     24751-
   2  01          31071    346607     30974      19173-     29424
   3  01           5201-    30974    359305      52728      31252-
   4  01            356-    19173-    52728     291616      53698
   5  01          24752-    29424     31253-     53698     313773
```

TABLE 1.4.3

```
          FA 1        I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
ROW  COL                              AFTER ITERATION 20
  1  01       41481     31359     5884-      1331    26722-
  2  01       31360    346820     31416    20901-    31259
  3  01        5883-    31416    358594     55144    34730-
  4  01        1329     20901-    55144    283688    64169
  5  01       26722-    31259     34730-    64169   301620
ROW  COL                              AFTER ITERATION 23
  1  01       41154     31764     6620-      3211    28904-
  2  01       31764    346449     32170    23473-    34046
  3  01        6618-    32169    357811     57883    38720-
  4  01        3209     23473-    57883    274489    76260
  5  01       28905-    34046     38720-    76260   287543
ROW  COL                              AFTER ITERATION 27
  1  01       40762     32344     7518-      5569    31626-
  2  01       32345    345541     33336    27220-    38151
  3  01        7518-    33336    356878     61355    43800-
  4  01        5568     27219-    61354    262663    91774
  5  01       31625-    38150     43800-    91774   269412
ROW  COL                              AFTER ITERATION 30
  1  01       40525     32729     8054-      7017    33297-
  2  01       32731    344837     34131    29703-    40901
  3  01        8054-    34130    356320     63517    46943-
  4  01        7017     29703-    63518    255255   101437
  5  01       33298-    40901     46943-   101437   258047
ROW  COL                              AFTER ITERATION 33
  1  01       40342     33049     8469-      8154    34612-
  2  01       33050    344242     34775    31725-    43155
  3  01        8469-    34776    355878     65227    49424-
  4  01        8156     31725-    65226    249372   109096
  5  01       34610-    43155     49424-   109096   249039
ROW  COL                              AFTER ITERATION 39
  1  01       40107     33488     9007-      9659    36347-
  2  01       33490    343442     35654    34468-    46242
  3  01        9008-    35653    355280     67521    52709-
  4  01        9659     34467-    67522    241509   119283
  5  01       36346-    46241     52710-   119282   237013
```

TABLE 1.4.4

```
        FA 1      I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES

ROW  COL                     AFTER ITERATION 40

 1   01        40081     33539      9070-      9828     36550-
 2   01        33540    343323     35743     34783-     46605
 3   01         9070-    35744    355204     67799      53085-
 4   01         9827     34782-    67799    240571     120455
 5   01        36549-    46604     53085-   120455     235611
ROW  COL                     AFTER ITERATION 41

 1   01        40054     33587      9129-      9990     36728-
 2   01        33588    343244     35853     35077-     46920
 3   01         9129-    35852    355174     68037      53432-
 4   01         9990     35077-    68038    239758     121537
 5   01        36728-    46920     53433-   121557     234342
ROW  COL                     AFTER ITERATION 42

 1   01        40035     33629      9174-     10127     36892-
 2   01        33631    343165     35947     35338-     47225
 3   01         9176-    35947    355096     68251      53729-
 4   01        10126     35338-    68250    238995     122485
 5   01        36893-    47225     53729-   122485     233210
```

```
                 FA 1                           CONVERGENCE TEST

               LOG DET                                                      CHANGE FROM
ITER-       I - F PRIME            SUM LOG                                    PREVIOUS
ATION       R INVERSE F            I - H SQ          DIFFERENCE               ITERATION

CENTROID     - 3.48896            - 8.42191            4.93295
    1        - 3.33270            - 8.36605            5.03335                0.10040
    2        - 3.35037            - 8.38894            5.03857                0.00522
    3        - 3.35846            - 8.39754            5.03908                0.00051
    4        - 3.36024            - 8.39946            5.03922                0.00014
    5        - 3.36145            - 8.40076            5.03931                0.00009
    6        - 3.36257            - 8.40197            5.03940                0.00009
    7        - 3.36382            - 8.40334            5.03952                0.00012
    8        - 3.36540            - 8.40506            5.03966                0.00014
    9        - 3.36751            - 8.40730            5.03979                0.00013
   10        - 3.37036            - 8.41034            5.03998                0.00019
   11        - 3.37420            - 8.41435            5.04015                0.00017
   12        - 3.37918            - 8.41955            5.04037                0.00022
   13        - 3.38568            - 8.42630            5.04062                0.00025
   15        - 3.40401            - 8.44524            5.04123                0.00061
   17        - 3.43106            - 8.47300            5.04194                0.00071
   20        - 3.49101            - 8.53416            5.04315                0.00121
   23        - 3.57468            - 8.61901            5.04433                0.00118
   27        - 3.71567            - 8.76132            5.04565                0.00132
   30        - 3.83611            - 8.88243            5.04632                0.00067
   33        - 3.96389            - 9.01067            5.04678                0.00046
   39        - 4.92304            - 9.27743            5.04739                0.00061
   40        - 4.27512            - 9.32255            5.04743                0.00004
   41        - 4.32048            - 9.36795            5.04747                0.00004
   42        - 4.36539            - 9.41298            5.04759                0.00012
   43        - 4.41091            - 9.45850            5.04759

DETERMINANT R EQUALS 0.000002834
```

TABLE 1.5.1

TABLE 1.5.2

FA 1 - - DETERMINATION OF THE NUMBER OF FACTORS

$N = 211$, $p = 24$, $|R| = 2.834 \times 10^{-6}$

Degrees of freedom $= p(p-1)/2 = 276$, $\chi^2_{.95}(276) = 315$ (approx.)

$\text{Log}_{10}|R| = 0.45250-6 = -5.54760$

$m_4 = N - k - (2p+11)/6 = 197.16667$, $m_5 = 196.16667$

	Centroid		Iteration 42	
	4 factors	5 factors	4 factors	5 factors
$\log_{10}\lvert I - F'R^{-1}F\rvert$	- 2.96222	- 3.48896	- 3.02993	- 4.36539
$\Sigma\log_{10}(1-h^2)$	- 7.61426	- 8.42191	- 7.61548	- 9.41298
Difference	4.65204	4.93295	4.58555	5.04759
$\log_{10}\lvert R\rvert$	- 5.54760	- 5.54760	- 5.54760	- 5.54760
$\log_{10}\lvert R(\underline{z}\lvert\underline{x})\rvert$	- 0.89556	- 0.61465	- 0.96205	- 0.50001
$\log_e\lvert R(\underline{z}\lvert\underline{x})\rvert$	- 2.06210	- 1.41528	- 2.21520	- 1.15132
$-m_k\log_e\lvert R(\underline{z}\lvert\underline{x})\rvert$	406.57738	277.63076	436.76360	225.85061
$\sqrt{2\chi^2} - \sqrt{2\nu-1}$	5.04	0.091	6.08	-2.22
Probability that k factors sufficient	< 0.00001	0.465	< 0.000001	0.987

```
FA 1    D SUB LAMBDA - - CHARACTERISTIC ROOTS OF  F PRIME F      6 DEC PLACES

                      FOR CENTROID SOLUTION                     ROW   COL

   9630727                                       7           1    01
              866140               1                         2    01
                        905894                               3    01
             1                  704020                       4    01
         7                                  597811           5    01

                      FOR FIFTH ITERATION                       ROW   COL

   9588764                                                   1    01
              807984                             7           2    01
                        878754                               3    01
                                690786                       4    01
             7                              572192           5    01

                  FOR FORTY-SECOND ITERATION                    ROW   COL

   9595613        4-                  5-         1           1    01
        4-    816184                             1           2    01
                        882068                               3    01
        5-                      699983           7           4    01
         1         1                   7    577740           5    01
```

TABLE 1.6.1

TABLE 1.7.1

```
FA 1     P MATRICES  - -  PRINCIPAL AXES SOLUTIONS      6 DECIMAL PLACES

ROW  COL             AFTER CENTROID

  1   01       758516    151298-    108537     161193     125872
  2   01       515166    294647     153516      66848-     35215
  3   01       661572    348259-    152533-      9995      76180
  4   01       479730    279229     104031-     61513     175637
  5   01       745045    135495-    252837      96019     148760-
  6   01       719170     39907-     40812     260429     240568-
  7   01       526757    128377-     26157-    191367      66612
  8   01       472896    268024-    160852-     94418-     20824
  9   01       718518     85536-    175886      54648     205676-
 10   01       704143     22444      25124     393914     274130
 11   01       696653    121711-    397304     312771-    195998
 12   01       504075    168648-    349725-     54648-     20472-
 13   01       555394    242648-    326591-    185709-     24876-
 14   01       624586     52360-    194238-    224683-     23239
 15   01       694394     40558-    284865      59923     207269-
 16   01       619329    172193      39139-     17012-    225700-
 17   01       521104    318437      99722-    121742-     86471-
 18   01       593500    215026      51875-     83733-    190245-
 19   01       615390     11475-    212890-    128060      41370-
 20   01       670382    312620     107857-     88369-     26864-
 21   01       752046     45398-    285256     331898-    192174
 22   01       729466    189585      53218-    175391     294664
 23   01       515686    163537     223141-    100306-    100771
 24   01       646081     18320-     44052-     47639-    166869-
```

TABLE 1.7.2

```
FA 1      P MATRICES  - -  PRINCIPAL AXES SOLUTIONS       6 DECIMAL PLACES

ROW  COL            AFTER FIFTH ITERATION

  1   01        758573    121971-    123513    163609    117723
  2   01        513862    302050     141675     71779-    31247
  3   01        663352    352456-    131943-    18893     78139
  4   01        479503    276432     108281-    63766    177298
  5   01        744101    120253-    258126     91352    160015-
  6   01        718233     32228-     42181    257098    247759-
  7   01        527111    125407-     13150-   195008     62624
  8   01        474487    274510-    148032-    87372-    25454
  9   01        717625     74310-    176489     49929    213680-
 10   01        703993     30357      41856    398394    265084
 11   01        696509    105526-    404899    314671-   187103
 12   01        505872    183044-    341506-    45808-    10558-
 13   01        557523    257344-    317387-   176425-    13796-
 14   01        625809     61646-    192763-   219151-    31272
 15   01        692964     24529-    283238     52587    218489-
 16   01        618447    172363      51955-    21021-   225064-
 17   01        520434    313995     116907-   124110-    81724-
 18   01        592704    213647      66773-    87424-   187961-
 19   01        615891     17323-    208961-   132298     38470-
 20   01        669838    308658     121844-    89373-    22821-
 21   01        752078     34342-    289342    332532-   187073
 22   01        729431    191234      47018-   179905    292003
 23   01        516319    153787     227066-    95607-   108391
 24   01        645990     18511-     46969-    48414-   166179-
```

TABLE 1.7.3

```
FA 1     P MATRICES  - -  PRINCIPAL AXES SOLUTIONS       6 DECIMAL PLACES

ROW  COL           AFTER FORTY-SECOND ITERATION

  1   01       753852    120951-   124834    170848    111523
  2   01       507541    254752     53995     65347-    11974
  3   01       654130    345121-   142784-    22975     15039-
  4   01       470235    269176    163799-    63007    133338
  5   01       742338     78940-   207091    202945    241464-
  6   01       718244     41080-    39486    227357    177619-
  7   01       518744    126905-    23181-   114160     74114
  8   01       469134    296804-   152975-    71103-     5880-
  9   01       717589     50813-   187781     20752    167729-
 10   01       708644      8914     12406    399312    307524
 11   01       693816     75927-   351243    242236-   116484
 12   01       498161    212639-   317811-    74816-    20598-
 13   01       557570    275973-   274120-   148895-    34703
 14   01       624984     69655-   206984-   177221-    34050-
 15   01       701009     23616    301683     90778    231878-
 16   01       614808    143363     43873-    40779-   158616-
 17   01       509285    293085    193626-    98395-   140099-
 18   01       586528    225236    114950-    63274-   181932-
 19   01       616067     37403-   135626-    17470     68639
 20   01       672218    293709    120395-    95242-    10172
 21   01       768232     25049-   389451    433454-   244037
 22   01       730689    183517     38084-   195658    295774
 23   01       523066    136121    199618-   110133-    83134
 24   01       643864      9490-    26449-    45506-   149326-
```

TABLE 1.8.1

```
FA 1     FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES

                             AFTER ÇENTROID

ROW COL
  1  01       770188    173400-     48926-     25873-    140966
  2  01       521460     23908      54450     271218     180199-
  3  01       656421     66825      86852-    277882-    260798
  4  01       468033     95391     176630-    306046      23377
  5  01       758122    224000-     45093     161782-    124882-
  6  01       704117    191692-    244273-    153397-    174454-
  7  01       524434    118680-    153484-     69745-    134750
  8  01       465344    152778      32808-    246790-    169767
  9  01       721799    136547-      1390     165226-    181299-
 10  01       707666    264917-    259289-    184991     234184
 11  01       745597     15390-    482279      33378      66103
 12  01       474577    265338     219594-    213290-    137498
 13  01       531606    336371     107174-    282425-    148686
 14  01       610337    314276      24291-     83854-     51935
 15  01       706168    205436-     66108     112329-    231651-
 16  01       598100     95587     145178-      1355     279360-
 17  01       499723    233409     111707-    183609     234094-
 18  01       573046    159237     109182-     50211     276034-
 19  01       589305     73345     291035-     65436-     25032
 20  01       649774    221112     140270-    209751     180384-
 21  01       789052     87694     397113      75446      49076
 22  01       726603     23381-    199251-    307751     159912
 23  01       495520    273792     145546-    141577      30885
 24  01       632242     90904      82178-    126436-    138126-
```

TABLE 1.8.2

```
FA 1     FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES

                             AFTER 5 ITERATIONS

ROW  COL
  1   01      .771386    126240-   168522     29819-     77014-
  2   01       522298    294515     27181    136974      49542
  3   01       655203    362344-   110501-    80486-     94124-
  4   01       467294    135138       596-   294219     174276-
  5   01       756878     35737    120731    282566-     12876
  6   01       697258     82823     61477    275011-    271520-
  7   01       523210    145216-    97195     40258-    172604-
  8   01       463897    271275-   192321-    63662-     29290-
  9   01       718720     80778     24376    275008-     22422-
 10   01       708561     98276-   326052    153614     292648-
 11   01       754814     37319-    68449     26353     466597
 12   01       469308    228258-   305934-     1122-    205667-
 13   01       527479    274336-   383286-    28140-     88959-
 14   01       608239     90200-   312959-    74147       8265-
 15   01       704367    152064     99287    284696-     36895
 16   01       591722    244034    164661-    89583-    145450-
 17   01       495386    292902    217775-   130235      98025-
 18   01       567452    264067    203188-    25232-    103533-
 19   01       583566     51565-   112222-     7646-    293955-
 20   01       645819    260922    190244-   170368     130106-
 21   01       796477      3123-    22914-    95037     390063
 22   01       727578     22263    133523    309768     211737-
 23   01       493201     42053    214434-   234388     129417-
 24   01       627554     57799    167610-   131768-     84801-
```

```
FA 1      FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES

                                 AFTER 42 ITERATIONS
ROW COL
 1   01     634078     133087-    150191      74771    438426-
 2   01     462102     250548      63388-     56470    215188-
 3   01     490442     325662-     86609-   161631-    433499-
 4   01     333003     174234     187952-    215794    344985-
 5   01     564137      42584     282697-    200846-   509269-
 6   01     489302      25012     146466     120444-   571278-
 7   01     399892     147931-     20528      33053    347782-
 8   01     368852     281566-    143874-    160681-   273538-
 9   01     619870      53031     153600     191042-   366335-
10   01     511429     121670-    136522     375665    565955-
11   01     817074      24876-    103752      51662-    33190-
12   01     331609     225033-    291396-    143025-   366653-
13   01     437288     288996-    298841-    148865-   315208-
14   01     507761      61286-    285214-    145515-   326753-
15   01     604272     156686     283179     193909-   370818-
16   01     473723     193651      92066-    125988-   375535-
17   01     355540     303026     263920-     68703-   345634-
18   01     426781     266924     169011-    121589-   385128-
19   01     474260      72893-    141703-     23549    391763-
20   01     547947     263648     236063-     52799    365435-
21   01     989867       1434-      1541-      4077    100303
22   01     581322      50994      41588-    356894    474513-
23   01     421111      79349     286247-     59627    286167-
24   01     511297      53970      60264-    177808-   374346-
```

TABLE 1.8.3

FA 1 - - FREQUENCY LIST OF DIFFERENCES AMONG P MATRICES

Amount of(±) Difference	Centroid vs 5th Iter.		Centroid vs 42nd Iter.		5th vs 42nd Iter.	
	Factor I	Others	Factor I	Others	Factor I	Others
0.00	24	47	13	13	12	12
0.01		43	10	21	11	24
0.02		6	1	17	1	19
0.03				16		12
0.04				5		6
0.05				7		6
0.06				5		4
0.07				2		5
0.08				3		3
0.09				2		1
0.10+				5		4
(max. diff.)				(0.111)		(0.115)

TABLE 1.9.1

```
          FA 111     R     CORRELATION MATRIX -- 6 DECIMAL PLACES

ROW  COL
  1  01      1000000    140100-    531100-    683200-    776200-    138600-    655900
  1  08        89800    147600-      93600     778600     393200     783100     313000
  1  15       421200     441100     676200     386900     164000     347100     222800
  1  22        49900-    186500-    171200-    215700-
  2  01       140100-   1000000     540400-    500100-    654100     290900-    352400
  2  08       172100     279100-    608300     191400-    449000      64000     700000
  2  15       677100     368800     396500     412400     412100     165600      16900
  2  22       224300-     75600-    210900-    269200-
  3  01       531100-    540400-   1000000     882500      96800     636000     555200-
  3  08       193600-    645300     288800-    289900-    440300-    500500-    680300-
  3  15       710000-    599400-    722900-    410000-    254800-    172000-     91900
  3  22       343500     280400     363300     506100
  4  01       683200-    500100-    882500    1000000     235000     417000     809500-
  4  08       396200-    422100     482400-    534800-    684000-    737300-    750300-
  4  15       801000-    608700-    814900-    603700-    419100-    351900-     93800-
  4  22       222300     206600     295500     409200
  5  01       776200-    654100      96800     235000    1000000      58000-    294300-
  5  08         6200-     52400-    289400     721800-     41200-    556000-    165600
  5  15        66800     124300-    278600-     88000-    104600     179300-    126100-
  5  22       109800-     85300      16000-     16000
  6  01       138600-    290900-    636000     417000      58000-   1000000     168900
  6  08        76900-    967100     282500     256700     129500      75100     204000-
  6  15       200700-    114200-    143000-    163500     204900     359400     584100
  6  22       355500     260300     314100     379900
  7  01       655900     352400     555200-    809500-    294300-    168900    1000000
  7  08       392900     143000     704500     758900     824100     863600     676900
  7  15       736000     588300     788400     754500     585200     609500     463100
  7  22        21900-     67600-    121400-    212300-
  8  01        89800     172100     193600-    396200-      6200-     76900-    392900
  8  08      1000000      79500-    349800     262200     527600     403800     326100
  8  15       268700      12400-    114900     292900     206200       3000-     63700-
  8  22        29800      40500      25300      30600-

DETERMINANT R EQUALS    .0000000000000008631
```

TABLE 2.1.1

TABLE 2.1.2

```
        FA 111     R     CORRELATION MATRIX -- 6 DECIMAL PLACES

ROW  COL
  9  01       147000-   279100-   645300    422100     52400-   967100    143000
  9  08        79500-  1000000    292500    255400    126700     63400    204700-
  9  15       190700-   110800-   138400-   174300    222400    383000    595400
  9  22       349400    281400    320300    387100
 10  01        93600    608300    288800-   482400-   289400    282500    704500
 10  08       349800    292500   1000000    418000    788100    484700    665900
 10  15       640400    467700    529700    633900    586700    454900    437400
 10  22         1000-    32000     39500-    98200-
 11  01       778600    191400-   289900-   534800-   721800-   256700    758900
 11  08       262200    255400    418000   1000000    589300    887200    309100
 11  15       388900    451100    639600    530700    313000    493900    432000
 11  22       101200     61300-     7200-    89200-
 12  01       393200    449000    440300-   684000-    41200-   129500    824100
 12  08       527600    126700    788100    589300   1000000    712000    696200
 12  15       673600    463000    644000    693900    536100    442500    360100
 12  22        22400-    16800-    83800-   165200-
 13  01       783100     64000    500500-   737300-   556000-    75100    863600
 13  08       403800     63400    484700    887200    712000   1000000    520500
 13  15       595300    530100    736000    626900    411400    495700    348000
 13  22         6300-    96300-   100000-   189900-
 14  01       313000    700000    680300-   750300-   165600    204000-   676900
 14  08       326100    204700-   665900    309100    696200    520500   1000000
 14  15       862000    600300    711300    623000    570700    340100    206100
 14  22       200200-   146800-   180200-   332000-
 15  01       421200    677100    710000-   801000-    66800    200700-   736000
 15  08       268700    190700-   640400    388900    673600    595300    662000
 15  15      1000000    596500    752600    660900    660500    451600    230800
 15  22       213500-   136900-   190800-   351100-
 16  01       441100    368800    599400-   608700-   124300-   114200-   588300
 16  08        12400-   110800-   467700    451100    463000    530100    600300
 16  15       596500   1000000    775500    490700    381700    419300    378000
 16  22       216800-   129500-   230700-   303800-
```

TABLE 2.1.3

```
          FA 111     R     CORRELATION MATRIX -- 6 DECIMAL PLACES

ROW  COL
 17   01      676200    396500    722900-   814900-   278600-   143000-   788400
 17   08      114900    138400-   529700    639600    644000    736000    711300
 17   15      752600    775500   1000000    653200    444700    531100    379400
 17   22      170400-   151300-   223400-   357000-
 18   01      386900    412400    410000-   603700-    88000-   163500    754500
 18   08      292900    174300    633900    530700    693900    626900    623000
 18   15      660900    490700    653200   1000000    577800    509600    448200
 18   22       32900     29200     23500    179500-
 19   01      164000    412100    254800-   419100-   104600    204900    585200
 19   08      206200    222400    586700    313000    536100    411400    570700
 19   15      660500    381700    444700    577800   1000000    503800    431100
 19   22        9700-    80000     65300    107900-
 20   01      347100    165600    172000-   351900-   179300-   359400    609500
 20   08        3000-   383000    454900    493900    442500    495700    340100
 20   15      451600    419300    531100    589600    503800   1000000    517900
 20   22       61100     89000     77900     51600-
 21   01      222800     16900     91900     93800-   126100-   584100    463100
 21   08       63700-   595400    437400    432000    360100    348000    206100
 21   15      230800    378000    379400    448200    431100    517900   1000000
 21   22      152700     45400    105200     73200
 22   01       49900-   224300-   343500    222300    109800-   355500     21900-
 22   08       29800    349400      1000-   101200     22400-     6300-   200200-
 22   15      213500-   216800-   170400-    32900      9700-    61100    152700
 22   22     1000000    442900    587700    625900
 23   01      186500-    75600-   280400    206600     65300    260300     67600-
 23   08       40500    281400     32000     61300-    16800-    96300-   146800-
 23   15      136900-   129500-   151300-    29200     80000     89000     45400
 23   22      442900   1000000    544600    592600
 24   01      171200-   210900-   363300    295500     16000-   314100    121400-
 24   08       25300    320300     39500-     7200-    83800-   100000-   180200-
 24   15      190800-   230700-   223400-    23500     65300     77900    105200
 24   22      587700    544600   1000000    429600
 25   01      215700-   269200-   506100    409200     16000    379900    212300-
 25   08       30600-   387100     98200-    89200-   165200-   189900-   332000-
 25   15      351100-   303800-   357000-   179500-   107900-    51600-    73200
 25   22      625900    592600    429600   1000000
```

TABLE 2.2.1

```
FA 111     F MATRIX FROM IMPROVED CENTROID SOLUTION          6 DECIMAL PLACES

FAC VAR        ELEMENT        FACTOR VAR        ELEMENT        FACTOR VAR        ELEMENT

 1  01         664839            3  01         584804-            5  01          92152-
 1  02         501307            3  02         680799             5  02         140431-
 1  03         805020-           3  03         108677             5  03         194771
 1  04         936529-           3  04          51725             5  04          46836
 1  05         213342-           3  05         820963             5  05          14404-
 1  06         221694-           3  06         244984             5  06         170970
 1  07         877057            3  07          87202             5  07          75734
 1  08         322134            3  08         193541             5  08         471346
 1  09         222764-           3  09         260605             5  09         143431
 1  10         615466            3  10         564334             5  10         146076
 1  11         642000            3  11         387000-            5  11         151000
 1  12         763690            3  12         302265             5  12         248758
 1  13         803664            3  13         232332-            5  13         139178
 1  14         806477            3  14         366214             5  14          88424-
 1  15         861479            3  15         322860             5  15         152147-
 1  16         709880            3  16          40221-            5  16         325719-
 1  17         890034            3  17          88281-            5  17         268304-
 1  18         698502            3  18         273772             5  18          71449-
 1  19         531094            3  19         452525             5  19         112507-
 1  20         475764            3  20         133715             5  20         215170-
 1  21         286438            3  21         146538             5  21         161345-
 1  22         298813-           3  22         113219             5  22         209478-
 1  23         285539-           3  23         247911             5  23         302238-
 1  24         331074-           3  24         193888             5  24         243267-
 1  25         489564-           3  25         148928             5  25         181936-
```

TABLE 2.2.2

```
FA 111    F MATRIX FROM IMPROVED CENTROID SOLUTION         6 DECIMAL PLACES

FAC VAR        ELEMENT         FACTOR VAR        ELEMENT         FACTOR VAR        ELEMENT

 2  01          241473              4  01         155682-             6  01         178197-
 2  02          396337-             4  02          33825-             6  02         107429
 2  03          406966              4  03         206043              6  03          47376-
 2  04          108705              4  04         271956              6  04          17097
 2  05          416461-             4  05         126486              6  05         194387
 2  06          801669              4  06         374953              6  06          80405-
 2  07          393087              4  07          53045-             6  07          55057-
 2  08           88458              4  08         448046-             6  08          45938
 2  09          808612              4  09         391131              6  09          67806-
 2  10          288229              4  10          91758              6  10         241178
 2  11          613000              4  11          35000-             6  11          47000
 2  12          309116              4  12         129916-             6  12         150847
 2  13          424982              4  13         151953-             6  13          18360
 2  14           99193-             4  14          76237-             6  14          63969
 2  15           66766-             4  15          66198-             6  15          95717-
 2  16           45509              4  16         229573              6  16         272590
 2  17          118348              4  17          63425              6  17         110014
 2  18          338094              4  18          10267-             6  18          88794-
 2  19          259855              4  19          81863              6  19         189756-
 2  20          472823              4  20         218580              6  20         151890-
 2  21          597013              4  21         425274              6  21          28084-
 2  22          548621              4  22         408311-             6  22          35001-
 2  23          422641              4  23         352682-             6  23         191115
 2  24          466337              4  24         325998-             6  24          85416-
 2  25          480789              4  25         333302-             6  25         183308
```

TABLE 2.2.3

```
FA 111    F MATRIX FROM IMPROVED CENTROID SOLUTION          6 DECIMAL PLACES

FAC VAR        ELEMENT         FACTOR VAR        ELEMENT

  7  01        151731-              9  01          46285
  7  02        135100-              9  02          61619-
  7  03         44119               9  03          71661
  7  04         57142               9  04          97227
  7  05         31379-              9  05          33190-
  7  06        117141-              9  06          13727
  7  07        111473-              9  07          88897-
  7  08         94947               9  08          24301-
  7  09         59081-              9  09          22342
  7  10         89175-              9  10          16038
  7  11         53000               9  11         118000
  7  12         37252-              9  12          57496-
  7  13         54125               9  13          46187
  7  14         14837-              9  14         122175
  7  15         71357               9  15         187560
  7  16         10932-              9  16          28595
  7  17         55253-              9  17          24562-
  7  18         69760               9  18         116758-
  7  19        191245               9  19         146115
  7  20         82518               9  20         155737-
  7  21        122080-              9  21          32853
  7  22        190969-              9  22          60219
  7  23        159880               9  23         151158-
  7  24        183473               9  24           4308-
  7  25        155961-              9  25         138592
```

TABLE 2.2.4

```
FA 111     F MATRIX FROM IMPROVED CENTROID SOLUTION          6 DECIMAL PLACES

FAC VAR       ELEMENT        FACTOR VAR       ELEMENT        FACTOR VAR       ELEMENT

  8  01        110354          10    1          15126
  8  02         27112          10    2        136367-
  8  03          9522          10    3          47608
  8  04         43726-         10    4          58436-
  8  05         15065-         10    5          23263-
  8  06         62460          10    6          60398-
  8  07         91238          10    7           5490-
  8  08         60781-         10    8         132530
  8  09         79976          10    9          44936-
  8  10         42039-         10   10         114509-
  8  11         36000-         10   11         129000-
  8  12         55789-         10   12          86150
  8  13         49523          10   13          38979-
  8  14         63065-         10   14          15991
  8  15        104533          10   15          42407-
  8  16         37300-         10   16           7990
  8  17         32188-         10   17          40571-
  8  18         73100-         10   18          36453-
  8  19         59550          10   19         102281
  8  20        110233          10   20         135512-
  8  21         97913-         10   21         133668
  8  22        179623-         10   22         106090-
  8  23        158347          10   23          26433
  8  24        225716-         10   24          53529-
  8  25        205353          10   25          86570
```

TABLE 2.2.5

```
FA 111              F MATRIX AFTER 12 ITERATIONS             6 DECIMAL PLACES

FAC VAR        ELEMENT        FACTOR VAR        ELEMENT        FACTOR VAR        ELEMENT
 1  01         665024              3  01        580329-             5  01         91759-
 1  02         502968              3  02        672001              5  02        132828-
 1  03         809032-             3  03        102862              5  03        183718
 1  04         934981-             3  04         55339              5  04         42938
 1  05         217558-             3  05        831359              5  05         14707-
 1  06         219483-             3  06        254790              5  06        160936
 1  07         876899              3  07         90760              5  07         82020
 1  08         335759              3  08        217268              5  08        546543
 1  09         222763-             3  09        257852              5  09        134659
 1  10         619112              3  10        563636              5  10        173268
 1  11         638720              3  11        387859-             5  11        152670
 1  12         755325              3  12        286645              5  12        206465
 1  13         798494              3  13        221956-             5  13        129686
 1  14         808148              3  14        365103              5  14        105734-
 1  15         865760              3  15        331712              5  15        158153-
 1  16         711615              3  16         35386-             5  16        312768-
 1  17         893442              3  17         88580-             5  17        279563-
 1  18         694147              3  18        260621              5  18         78075-
 1  19         526971              3  19        441894              5  19         88594-
 1  20         478890              3  20        138189              5  20        194013-
 1  21         283535              3  21        150592              5  21        122838-
 1  22         306850-             3  22        137335              5  22        242764-
 1  23         295164-             3  23        254732              5  23        345972-
 1  24         329120-             3  24        189398              5  24        232550-
 1  25         484850-             3  25        143727              5  25        163776-
```

TABLE 2.2.6

```
FA 111              F MATRIX AFTER 12 ITERATIONS              6 DECIMAL PLACES

FAC VAR       ELEMENT        FACTOR VAR       ELEMENT        FACTOR VAR       ELEMENT

 2  01        239642            4  01         144260-           6  01         142017-
 2  02        391719-           4  02          30224-           6  02         130016
 2  03        396497            4  03         208168            6  03          41239-
 2  04        110567            4  04         284502            6  04          18439
 2  05        425487-           4  05         121244            6  05         214826
 2  06        812164            4  06         399918            6  06          99774-
 2  07        393278            4  07          58943-           6  07          64694-
 2  08        111706            4  08         484782-           6  08          74739
 2  09        805936            4  09         373264            6  09          72380-
 2  10        288327            4  10          79557            6  10         274179
 2  11        613443            4  11          40972-           6  11          41927
 2  12        901834            4  12         111142-           6  12         137760
 2  13        421639            4  13         139397-           6  13           2006-
 2  14         95559-           4  14          46430-           6  14          55112
 2  15         67329-           4  15          66127-           6  15         119448-
 2  16         50842            4  16         215772            6  16         231152
 2  17        124160            4  17          77637            6  17         117955
 2  18        333187            4  18            465            6  18          80579-
 2  19        247166            4  19          51610            6  19         182864-
 2  20        459809            4  20         202833            6  20         148264-
 2  21        575231            4  21         395806            6  21          17090-
 2  22        568549            4  22         463044-           6  22          82219-
 2  23        448665            4  23         377802-           6  23         243378
 2  24        458879            4  24         295669-           6  24          74927-
 2  25        471300            4  25         323969-           6  25         152641
```

TABLE 2.2.7

```
FA 111            F MATRIX AFTER 12 ITERATIONS           6 DECIMAL PLACES

FAC VAR       ELEMENT          FACTOR VAR       ELEMENT

 7  01        104461-              9  01          44368
 7  02        106035-              9  02          52043-
 7  03         34878               9  03          85018
 7  04         69544               9  04         104399
 7  05         50563-              9  05          50260-
 7  06        108744-              9  06            645-
 7  07        125873-              9  07         106069-
 7  08        193585               9  08          49072-
 7  09         49190-              9  09          21890
 7  10         86661-              9  10          17942
 7  11         42609               9  11         115044
 7  12         39191-              9  12          23185-
 7  13         19485               9  13          58369
 7  14          8989               9  14         117049
 7  15         97155               9  15         223391
 7  16         11095-              9  16          16457-
 7  17         33973-              9  17            503-
 7  18         67858               9  18         101988-
 7  19        170307               9  19          60029
 7  20         60473               9  20         130456-
 7  21        105123-              9  21           3829-
 7  22        293378-              9  22         115724
 7  23        244223               9  23         183049-
 7  24        185317               9  24          22897-
 7  25        148556-              9  25         111145
```

TABLE 2.2.8

FA 111 F MATRIX AFTER 12 ITERATIONS 6 DECIMAL PLACES

FAC	VAR	ELEMENT	FACTOR	VAR	ELEMENT
8	01	119349	10	1	2294
8	02	20129	10	2	88769-
8	03	7943	10	3	34832
8	04	49761-	10	4	23756-
8	05	30949-	10	5	38028-
8	06	83151	10	6	8895-
8	07	88878	10	7	9578-
8	08	136479-	10	8	344388
8	09	72676	10	9	28871-
8	10	23112-	10	10	206053-
8	11	37670-	10	11	147396-
8	12	11461-	10	12	33862
8	13	17475	10	13	251-
8	14	20103-	10	14	64312
8	15	114463	10	15	46088-
8	16	61212-	10	16	9204
8	17	41708-	10	17	5063
8	18	86129-	10	18	93485-
8	19	73712	10	19	55108-
8	20	40632	10	20	119543-
8	21	61008-	10	21	35703
8	22	305262-	10	22	74014-
8	23	270878	10	23	8236
8	24	160528-	10	24	69575-
8	25	141739	10	25	59003

```
                FA 111   COMMUNALITIES      6 DECIMAL PLACES

VAR    1         2         3         4         5         6         7    ITERATION

    960461    942352    946515    992907    901861    945015    987740   AM 557
    944370    945665    917562    986070    949782    949584    971615   AM 561
    925116    944435    920393    988465    974290    973325    977338       1
    918745    932935    917594    990829    985507    981840    980640       2
    911517    918069    912840    991046    988030    986254    981469       4
    910842    917280    912232    991126    988013    986923    981375       5
    910662    916896    911877    991234    988025    987265    981281       6
    910679    916589    911622    991339    988057    987399    981207       7
    911695    915958    911134    991449    988158    987371    981246      10   P
    911695    915958    911134    991449    988158    987371    981246      10   C
    912341    915833    910999    991388    988168    987329    981319      11
    912997    915704    910831    991338    988082    987289    981329      12

VAR    8         9        10        11        12        13        14

    529805    959481    905112   1021561    907606    967539    856892   AM 557
    604940    961951    891361   1000338    886632    932296    831460   AM 561
    704497    959075    940761    982051    839909    909414    837828       1
    775702    955675    953801    990915    831887    905714    846710       2
    843977    953616    950426    996115    824595    905560    842746       4
    856723    953333    949166    997218    822980    905518    839016       5
    865506    953113    948335    997981    822022    905434    836342       6
    872617    952939    947726    998517    821392    905287    834359       7
    884550    952683    946735    999397    820771    905113    831697      10   P
    884550    952683    946735    999397    820770    904959    831010      10   C
    887468    952674    946584    999557    820846    905003    830471      11
    888839    952686    946467    999650    821095    904983    830229      12
```

TABLE 2.3.1

```
                FA 111     COMMUNALITIES      6 DECIMAL PLACES

VAR   15          16          17          18          19          20          21  ITERATION

    930820      741794      926476      735089      654366      631697      692168  AM 557
    940531      743113      908411      715430      681661      646517      711060  AM 561
    948287      746832      904581      698916      643930      624732      664167       1
    957630      751495      906298      691657      627891      607421      636536       2
    968401      740184      912163      688304      621572      594617      620156       4
    971166      733698      914367      687902-     619900      590777      619149       5
    973782      728662      916125      688160      618352      588331      619196       6
    976020      724780      917391      688790      617200      587288      619707       7
    979332      718696      919664      691218      616501      588096      620824      10  P
    980671      716368      920758      692845      616869      590365      621359      10  C
    981704      713842      921383      694510      617862      593286      621888      11
    982328      712271      922517      696912      619090      597238      622067      12

VAR   22          23          24          25
    661465      587964      609074      734275                                      AM 557
    698519      648058      624017      763999                                      AM 561
    723015      684483      631879      775451                                           1
    744575      698500      630923      787396                                           2
    797306      715756      625687      793504                                           4
    817808      724372      620314      789769                                           5
    835274      734774      614271      782297                                           6
    850839      747104      607820      771656                                           7
    878574      778725      593238      742550                                          10  P
    891005      797923      585028      725319                                          10  C
    903229      819097      576040      707748                                          11
    914496      841569      567352      690932                                          12
```

TABLE 2.3.2

```
ROW  COL   FA 111    I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
                                ITERATION 1
 1   01      10995      5128-      1138-    21575      4639     1181     12121
 1   08     13510-      6005      23428
 2   01      5129-     29091       8534    24008-    30163-    14468     18776
 2   08     21369-    20204-      20165
 3   01      1138-      8534      29951    25827-      966    36612-     19893
 3   08     23947-     14115      14168
 4   01      21574    24007-     25829-   117314    10214-    22361      7105
 4   08        808    14744-      36832
 5   01       4640    30163-        966    10214-   167484    15244-       309
 5   08      30104     16573          3
 6   01       1179     14467     36611-    22361    15244-   260902     12644
 6   08      94320     9997-      90535
 7   01      12122     18776      19893     7105       308    12645    340740
 7   08      53411   103873-      15151
 8   01     13509-    21370-     23947-      807     30104    94320     53411
 8   08     448792     16202     70615-
 9   01       6006    20203-      14114    14744-    16572     9998-  103873-
 9   08      16202    327846      82618
10   01      23427     20166      14168    36833         2    90535     15151
10   08     70614-     82619     438628

                                ITERATION 2
 1   01       9964     5196-      1497-    21657      2211      208      9756
 1   08     13651-      4411      23075
 2   01      5197-     26286       9345    22195-    32500-    13527     19287
 2   08     19686-    19909-      20564
 3   01      1496-      9346      24984    24559-     473-    34771-     21550
 3   08     24492-     10346       1675
 4   01      21656    22194-     24561-   107607     4930-    25252     11246
 4   08      6184-    13851-      41773
 5   01       2211    32499-       473-    4929-    147595    23754-   10694-
 5   08      36339     22163      4708-
 6   01        207     13528     34771-    25252    23754-   231427      7688
 6   08     101109     5239-     105204
 7   01       9757     19287      21549    11246    10694-     7688    324538
 7   08      51536   115900-       4003
 8   01     13650-    19685-     24492-    6184-    36339    101109     51536
 8   08     428626     15297     66228-
 9   01       4413    19908-      10346    13851-    22162     5239-  115900-
 9   08      15296    281498      85226
10   01      23075     20564       1675    41773     4708-   105204      4002
10   08     66228-     85226     389303
```

TABLE 2.4.1

```
              FA 111    I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
ROW  COL                           AFTER ITERATION 4

  1  01       9265     4816-     1545-    20638       582       143      5854
  1  08     15587-     2438     22656
  2  01      4817-    24066      8538    19619-    32517-    13154     21614
  2  08     14330-    19692-    20319
  3  01      1543-     8539     21860    22344-      410-    30187-    21031
  3  08     22382-     3638      8496-
  4  01     20638     19617-    22346-   97025      4939-    23754      9363
  4  08     18565-     8221-    45584
  5  01       584     32516-      410-    4940-   133340     30776-    22964-
  5  08     37460     23957     14924-
  6  01       141     13154     30186-   23754     30776-   212524      1418-
  6  08     97564      7831    112696
  7  01      5853     21614     21030     9365     22965-     1418-   298809
  7  08     32549    124734-     3174-
  6  01     15587-    14330-    22382-   18565-    37460     97564     32549
  8  08    387833     19020     57002-
  9  01      2439     19691-     3639     8220-    23957      7831    124735-
  9  08     19020    256978     91345
 10  01     22656     20320      8495-   45585     14923-   112696      3174-
 10  08     57002-    91346    343756

                                   ITERATION 7

  1  01       8784     4296-     1409-   19760       833-      770-     4120
  1  08     18697-       83     21707
  2  01      4297-    22147      7668    16799-    31247-    13642     22375
  2  08      8322-    18424-    21118
  3  01      1407-     7668     19374    19261-     1123     25975-    22207
  3  08     19633-     1980-    12868-
  4  01     19760     16798-    19262-   89521      8081-    17642      6577
  4  08     34109-     6644-    44929
  5  01       831-    31246-     1123     8081-   123473     37012-    29770-
  5  08     34340     23694     22846-
  6  01       771-    13642     25975-   17642     37012-   203622     18760-
  6  08     94898     28377    119165
  7  01      4121     22375     22206     6578     29771-    18760-   275089
  7  08       595-   129026-    13611-
  8  01     18696-     8322-    19632-   34109-    34341     94898       595-
  8  08    351871     33885     49719-
  9  01        82     18425-     1981-    6644-    23694     28377    129026-
  9  08     33885    239929     98650
 10  01     21706     21119     12868-   44929     22846-   119165     13611-
 10  08     49719-    98650    319897
```

TABLE 2.4.2

```
          FA 111    I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
ROW  COL                    ITERATION 10  PARTIAL

  1  01       8507     3846-     1229-    19139     1827-     1407-      4062
  1  08     20375-     1694-    20645
  2  01      3847-    20713       6947    14620-   29459-    13754      21045
  2  08      5732-    16408-    21949
  3  01      1230-     6947     18049    17088-     2213     24230-     22592
  3  08     17961-     4637-    14386-
  4  01     19139     14618-    17091-    85111    10854-    13062       6974
  4  08     43102-     8235-    43214
  5  01      1828-    29456-     2213    10854-   117161     38722-    30556-
  5  08     33605     21979     27170-
  6  01      1407-    13754     24230-   13062     38722-   200565     31672-
  6  08     93708     42908    123162
  7  01      4061     21045     22592     6974     30557-    31672-   259882
  7  08     23853-   128317-    19993-
  8  01     20375-     5732-    17961-   43102-    33605     93708     23853-
  8  08    332416     47777     42701-
  9  01      1694-    16408-     4638-    8234-    21978     42907    128317-
  9  08     47777    231357    102009
 10  01     20646     21948     14385-   43215     27170-   123162     19993-
 10  08     42702-   102010    310306

                            ITERATION 10  COMPLETE

  1  01       8334     3509-     1041-    18539     2303-     1619-      4519
  1  08     20996-     2666-    20090
  2  01      3508-    19864      6622    13338-    28017-    13387      20056
  2  08      5411-    15089-    22806
  3  01      1040-     6623     17650    16115-     3092     23866-     23363
  3  08     18152-     6038-    14495-
  4  01     18537     13337-    16117-   82453     13073-    11798       7603
  4  08     45329-     9271-    41124
  5  01      2301-    28017-     3092    13074-   114165     38700-    29140-
  5  08     32867     20801     27338-
  6  01      1620-    13387     23866-   11798     38700-   199236     40105-
  6  08     92779     51338    125746
  7  01      4519     20057     23364     7604     29141-    40105-   253717
  7  08     39207-   127975-    21894-
  6  01     20996-     5411-    18153-   45329-    32867     92779     39207-
  8  08    324543     56374     40102-
  9  01      2665-    15089-     6099-    9271-    20801     51337    127975-
  9  08     56374    228220    104781
 10  01     20091     22806     14495-   41124     27339-   125746     21894-
 10  C8     40102-   104781    310125
```

TABLE 2.4.3

```
            FA 111    I-F PRIME  R INVERSE  F   MATRICES   6 DECIMAL PLACES
ROW  COL                             ITERATION 11

  1  01     8142     3093-      882-     17911     3072-     1546-     4852
  1  08    21272-    3489-    19672
  2  01     3093-   18825      6184      11678-   26454-    12784     18828
  2  08     5452-   13765-    23365
  3  01      881-    6184     17169      15026-    3853     23487-    23698
  3  08    17759-    6938-    14805-
  4  01    17911    11676-    15026-     79573    15755-    11034      8581
  4  08    47745-   10470-    40048
  5  01     3071-   26454-     3853      15756-  110777     38212-    28119-
  5  08    33932    19470     29026-
  6  01     1545-   12784     23487-     11034    38213-   197006     47582-
  6  08    89817    58290    127424
  7  01     4854    18828     23699       8581    28120-    47582-   245960
  7  08    52510-  125826-    24065-
  8  01    21272-    5452-    17758-     47745-   33932     89818     52510-
  8  08   313966    64195     37395-
  9  01     3489-   13766-     6939-     10470-   19470     58290    125826-
  9  08    64195   224706    106287
 10  01    19671    23365     14805-     40048    29026-   127424     24066-
 10  08    37395-  106287    307605

                                     ITERATION 12

  1  01     7907     2624-      666-     17062     3809-     1380-     5284
  1  08    21160-    4226-    19166
  2  01     2623-   17648      5717       9861-   24506-    11818     17428
  2  08     6048-   12482-    23887
  3  01      666-    5717     16774      13910-    4602     23186-    23932
  3  08    17461-    7562-    14894-
  4  01    17059     9859-    13911-     76079    18486-    10873      9615
  4  08    48634-   11648-    38603
  5  01     3807-   24507-     4601      18487-  106794     36591-    26440-
  5  08    34975    17848     29551-
  6  01     1381-   11819     23186-     10873    36591-   193813     54589-
  6  08    85175    64242    128398
  7  01     5285    17429     23933       9614    26441-    54589-   237673
  7  08    65027-  122840-    25421-
  8  01    21160-    6048-    17461-     48633-   34975     85175     65027-
  8  08   302201    71768     35416-
  9  01     4225-   12482-     7563-     11649-   17848     64242    122840-
  9  08    71768   221651    107215
 10  01    19166    23887     14894-     38604    29552-   128397     25421-
 10  08    35416-  107215    306344
```

TABLE 2.4.4

TABLE 2.5.1

```
FA 111    LOG DET  I  MINUS  F PRIME  R INVERSE  F    6 DECIMAL PLACES
          ALL ELEMENTS ARE NEGATIVE

1 FACTOR  2 FACTORS 3 FACTORS 4 FACTORS 5 FACTORS             ITERATION
  1958804   3532309   5094020   6361980   7285897                     1
  2001566   3629096   5293678   6674637   7701449                     2
  2033154   3699465   5424684   6889732   8011998                     4
  2056307   3754301   5531143   7033453   8234823                     7
  2070223   3792064   5595667   7110954   8350916                    10   P
  2079146   3814659   5626040   7144368   8395986                    10   C
  2089268   3842524   5662541   7183513   8449915                    11
  2101988   3877273   5703592   7227123   8504727                    12

6 FACTORS 7 FACTORS 8 FACTORS 9 FACTORS 10 FACTORS            ITERATION
  8008568   8541579   8988569   9599983  10163726                     1
  8529857   9095724   9640179  10375718  11078686                     2
  8870793   9478189  10079919  10855344  11781762                     4
  9095590   9752065  10406852  11228923  12475378                     7
  9211352   9903127  10594004  11441881  12909101                    10   P
  9257102   9969148  10675576  11533725  13127219                    10   C
  9312770  10048410  10774781  11642432  13362742                    11
  9369824  10131354  10876789  11751805  13568649                    12

FA 111    SUM LOG  ONE  MINUS  H SQUARE                 6 DECIMAL PLACES
          ALL ELEMENTS ARE NEGATIVE

1 FACTOR  2 FACTORS 3 FACTORS 4 FACTORS 5 FACTORS             ITERATION
  6609660  10649183  15064961  17816460  19778070                     1
  6620834  10663017  15108024  17899578  19938275                     2
  6625916  10670820  15104363  17911477  20003699                     4
  6627337  10673433  15108904  17931890  20052945                     7
  6627454  10671614  15109126  17944616  20079455                    10   P
  6627917  10671703  15109409  17949408  20090283                    10   C
  6627975  10671050  15109637  17955555  20102538                    11
  6628278  10670153  15109501  17961646  20116546                    12

6 FACTORS 7 FACTORS 8 FACTORS 9 FACTORS 10 FACTORS            ITERATION
 21125181  22030019  22737768  23827341  24658711                     1
 21440362  22430236  23222612  24522761  25681413                     2
 21543259  22594647  23484543  24908397  26435533                     4
 21596793  22688997  23653995  25177900  27138033                     7
 21626498  22738046  23751519  25340569  27574262                    10   P
 21642378  22764341  23803558  25423803  27794370                    10   C
 21658674  22791894  23864900  25516950  28030600                    11
 21678656  22824899  23937424  25617636  28236971                    12
```

```
             FA 111                       CONVERGENCE TEST

          LOG DET                                                     CHANGE FROM
ITER-     I - F PRIME           SUM LOG                                  PREVIOUS
ATION     R INVERSE F           1 - H SQ          DIFFERENCE            ITERATION

  1        - 10.16373           - 24.65871         14.49498
  2        - 11.07869           - 25.68141         14.60272              0.10774
  4        - 11.78176           - 26.43553         14.65377              0.05105
  7        - 12.47538           - 27.13803         14.66265              0.00888
 10 PART   - 12.90910           - 27.57426         14.66516              0.00251
 10 COMPL  - 13.12722           - 27.79437         14.66715              0.00199
 11        - 13.36274           - 28.03060         14.66786              0.00071
 12        - 13.56865           - 28.23697         14.66832              0.00046
DETERMINANT R EQUALS 0.0000000000000008631
```

TABLE 2.5.2

TABLE 2.5.3

FA 111 - DETERMINATION OF THE NUMBER OF FACTORS

$N = 350,\ p = 25,\ |R| = 8.631 \times 10^{-16}$

Degrees of freecom = $p(p-1)/2 = 300,\ \chi^2_{.95}\,(300) = 342$ (approx.)

$\log_{10}|R| = 0.93607-16 = -15.06393$

$m_9 = N - k - (2p+11)/6 = 330.83333,\ m_{10} = 329.83333$

	Iteration 2		Iteration 12	
	9 factors	10 factors	9 factors	10 factors
$\log_{10}\lvert I - F'R^{-1}F\rvert$	- 10.37572	- 11.07869	- 11.75181	- 13.56865
$\Sigma \log_{10}(1-h^2)$	- 24.52276	- 25.68141	- 25.61763	- 28.23697
Difference	14.14704	14.60272	13.86582	14.66832
$\log_{10}\lvert R\rvert$	- 15.06393	- 15.06393	- 15.06393	- 15.06393
$\log_{10}\ \lvert R(\underline{z}\vert\underline{x})\rvert$	- 0.91689	- 0.46121	- 1.19811	- 0.39561
$\log_e\ \lvert R(\underline{z}\vert\underline{x})\rvert$	- 2.11122	- 1.06198	- 2.75875	- 0.91093
$-m_k\log_e\ \lvert R(\underline{z}\vert\underline{x})\rvert$	698.46195	350.27640	912.68645	300.45508
$\sqrt{2\chi^2}\ \ -\sqrt{2v-1}$	12.90	1.99	18.25	0.04
Probability that k factors sufficient	< 0.000001	0.023	< 0.000001	0.48

```
FA 111   D SUB LAMBDA - - CHARACTERISTIC ROOTS OF  F PRIME F      6 DEC PLACES
                                 FOR CENTROID SOLUTION               ROW   COL
9770936                                                               1    01
                                                                      1    08
     4401190                     4-                                   2    01
                                                                      2    08
             2976753                                                  3    01
  6-                                                                  3    08
         4-           1492969                                         4    01
                                                                      4    08
                                  980807                              5    01
                                                                      5    08
                                          234973                      6    01
                                                                      6    08
                                                  409290              7    01
               5                                                      7    08
               6-                                                     8    01
249607   9                                                            8    08
                                                                      9    01
      9  206985                                                       9    08
                                                                 5   10    01
              129526                                                 10    08

                                 FOR TWELTH ITERATION                ROW   COL
9769518                                                               1    01
                                                                      1    08
     4396249                                                          2    01
                                                                      2    08
             2978742                                                  3    01
                                                                      3    08
                      1582914                 1                       4    01
                                                                      4    08
                                 1094647                              5    01
  3-                                                                  5    08
                      1           383008                              6    01
  1                                                                   6    08
                                                  438875              7    01
                                                                      7    08
                                   3-         1                       8    01
230121                                                                8    08
                                                                      9    01
     175330                                                           9    08
                                                                     10    01
             148693                                                  10    08
```

TABLE 2.6.1

TABLE 2.7.1

```
FA 111   P MATRICES  - -  PRINCIPAL AXES SOLUTIONS      6 DECIMAL PLACES

ROW  COL                          AFTER CENTROID

 1   01      646311     38173   685923-    64467-    99228-    79641-     2199
 1   08      180759     31055-   26769
 2   01      498669    387687-  706842     49648-    89590-    84243-    64399-
 2   08       76845    120709-   69691-
 3   01      725155-   574267   101310    138781    157602     24788     37433
 3   08        6728     74061     4618
 4   01      903343-   304451   124223    204188      9498-     9020-     5001-
 4   08       67518-    86102    89765-
 5   01      203429-   254753-  913292     53442     18845      4704     77022-
 5   08        4684-    10495-    9943-
 6   01       74160-   895297   152924    301632    126274     40337-    27234-
 6   08       76520     58182-   24581-
 7   01      933634    238476    48081-    24884-   104805       740-    13239-
 7   08      105571    126365-   43196
 8   01      338901      7543    88747    406619-   551958     42323     48833
 8   08       41330-    28784    73787
 9   01       72740-   903926   171279    307476    100241     11038      4029-
 9   08       63228     27009-   28021-
10   01      701771    249342   501880     44426    179079     27081-   194280-
10   08       64689-     3696-   90016-
11   01      694378    430759   511852-     5162-   136774      3744     70169-
11   08      110153-    82000   162388-
12   01      822891    185364   192359    131058-   309313     27628    108127-
12   08       59732-     4648-   96051
```

```
FA 111   P MATRICES  - -  PRINCIPAL AXES SOLUTIONS      6 DECIMAL PLACES

ROW  COL                    AFTER CENTROID
13   01      837370    227856    362113-   107748-   157531     63739    16133-
13   08       11994-    35445     76903-
14   01      812699    198352-   320185     74176-    25882-    81305-     6287
14   08         600-   127645      1436
15   01      869137    171105-   249006     48020-    99668-    19616-   173023
15   08      142751    140885    105182-
16   01      709598     78687-    14924-   192504    327207-    84283    225869-
16   08      135697-   114168     19426
17   01      895146     53665-   124763-    65639    255510-      359    117224-
17   08       72148-     2897-    11380
18   01      766821    238723    164505     33742-    30396-     5000-   144192
18   08       92290-   102109-    39267
19   01      600805    232417    348102     40034     61970-    50868    313639
19   08       79688    177186     36842
20   01      559980    417149     45262    176157    228866-   102604    162672
20   08        5847    182238-    56841-
21   01      387530    588275     82024    348546    175881-    81711-    48944-
21   08       39478-    76554    197928
22   01      170760-   562329     17248    483046-   169627-   291331-    56812-
22   08       22932-    42730-    16965-
23   01      165639-   439292    203489    481274-   254065-   290294     41674-
23   08       15145-    61242-     8705
24   01      216668-   497756    114627    429322-   193698-    76969-   203232
24   08      216154-    17638      8591-
25   01      356105-   535517     98904    431960-   151297-    52298    227455-
25   08      236105    116464     12010
```

TABLE 2.7.2

```
FA 111   P MATRICES  - -  PRINCIPAL AXES SOLUTIONS      6 DECIMAL PLACES

ROW  COL                       AFTER TWELTH ITERATION

 1   01       645472     37588     674813-     60499-    111969-     12619-     17819-
 1   08       144410     37126-     25883
 2   01       499419    378355-    702742      34039-    102084-     24752      86046-
 2   08        70139     54460-     42473-
 3   01       731891-   557380      93802     142933     163139      23949-     25867
 3   08         6621     74900      41877
 4   01       902955-   299132     122966     216857       8607       2594-      9703
 4   08        69939-   138095      15693
 5   01       206902-   260099-    929377      58618      12182      76[illegible]0      62744-
 5   08         9795-    13741-     13091-
 6   01        74202-   902736     151316     337966     132879       6353-      4654-
 6   08        80702     62378-     40136
 7   01       932964    243997      45315-     22558-     96400       7936      16775-
 7   08       109478    162936-     25282-
 8   01       354289     17916     107002     498512-    676834      22929-     91118
 8   08       135074-    32378-    129643
 9   01        76782-   894985     161015     322857     108762       4330-     26496
 9   08        48365      8392-     25912
10   01       707271    254503     508925      56378     168795     198894     116691-
10   08        13420     92324     170022-
11   01       692081    433037     511065-      6583-    122644      82553      28229-
11   08        38031-   171818     135285-
12   01       814425    185706     183612     114147-    252023     108064      34421-
12   08         2705     13818-       524-
```

TABLE 2.7.3

TABLE 2.7.4

```
FA 111   P MATRICES  - -  PRINCIPAL AXES SOLUTIONS       6 DECIMAL PLACES

ROW  COL                          AFTER TWELTH ITERATION
 13   01        832242    230374    350780-    101002-    144987     30232      9504
 13   08         18597     58582     15232-
 14   01        813582    190693-   316988      51007-     37376-    48370-    34896-
 14   08          4669     88625    126628
 15   01        874615    166005-   291626      38902-    113966-   223110-    59916
 15   08        147944    182683     58079
 16   01        709619     75372-    20142-    182355     290741-   209114      7617-
 16   08        170837-    48918     97342
 17   01        896649     47115-   126182-     74469     252563-   113739     27015-
 17   08        102178-    19460     81090
 18   01        760053    238935    153843      19457-     35427-   116576-    37353
 18   08        118650-    82164-    32018-
 19   01        597184    220616    338312      31051      49799-   260350-   150580
 19   08         41429     60943      1793-
 20   01        559034    405762     47063     180151     200613-   113392-   117307
 20   08         46233-   104658-    73836-
 21   01        378451    562357     76613     333462     114723-    27681     66365-
 21   08         90735-    47423-   129286
 22   01        176363-   598673     24511     516456-    217197-   117404-   440948-
 22   08         38653-    27590-     2128-
 23   01        168052-   467596    229181     500940-    306082-   206874    390837
 23   08         27894     25183-    23826-
 24   01        217597-   501326    117301     383990-    193369-   169042-    52258
 24   08        179757-    37794     70915-
 25   01        368876-   530490     96916     400333-    147697-   165006     61103-
 25   08        189686     49027    112355
```

```
FA 111   FE MATRIX  - -  LAWLEYS SOLUTION  - -  6 DECIMAL PLACES

                  F MATRIX FROM CENTROID OPERATION

ROW COL
  1  01    799450   210664-  364573-    60891   328656-    25531     4181-
  1  08     30734    23579   118593-
  2  01    128482-  797490-  494765     42520   133268     79083-   51235-
  2  08     21420-   72357   119301-
  3  01    352427-  851267   216088    100775-   13277-    46398    73200
  3  08     16630    59977-   17089
  4  01    601681-  777856    48855     70984    94762      3190-   15978
  4  08       770    43759    20105-
  5  01    685905-  301385-  560729      1074   267923     22125    33744-
  5  08     12798-    2156    22021-
  6  01    224009   683656   628678     31663   171995-    16185-   17571-
  6  08     12689    62809-   38980-
  7  01    809412   398108-  338612     50248-  178025-    15806-   21059-
  7  08     16730-   89843-    7559-
  8  01    282364   242380-  118246    542595-  209349     33935   143206
  8  08     59259   292917-   57211
  9  01    218603   685987   642735     46416   163495-     7651     9334
  9  08     18240-   33174-    3614
 10  01    433438   340526-  725973     35856-  213689     23394   101746-
 10  08     23610    32688-   32016-
 11  01    995376    82405    16482-     1908    27280      1543-      34-
 11  08       730-    4069     1989-
 12  01    637063   411534-  452654    221126-  100625     72770    50536-
 12  08     69622   163939-   89071
```

TABLE 2.8.1

TABLE 2.8.2

```
FA 111   FE MATRIX - -   LAWLEYS SOLUTION   - -   6 DECIMAL PLACES

                      F MATRIX FROM CENTROID OPERATION
ROW COL
13   01        929671   228236-    27438    84020-    20854-    33167     46166
13   08         44796-   46113-    17478
14   01        384951   722132-   339273    43044    112269     57240     82573
14   08        118343    89865      7693
15   01        455882   724981-   314248   106449       517-    39665    268458
15   08          6145-  145444     46768-
16   01        497087   445633-   129566   372301     30488    155153    219486-
16   08          5069-  226962    131611
17   01        691711   543219-    93304   243467     76451-    23397    153377-
17   08         13893   176380     75235
18   01        574256   358191-   431995    19058-    97851-   147843-    51345
18   08         21180    67577    176900
19   01        346355   302294-   534314    53903    104318-    54914    352233
19   08         32686   132290    159329
20   01        516455    63737-   408963   190002    271094-   166270-    28265
20   08        167727-  152182    138379
21   01        437358   188660    506757   273828    242335-    93549     99773-
21   08        195241   106310    159837
22   01         72955   380928    151732   474641-   142549-   164233-    98163-
22   08        210513   430797    113808-
23   01         75984-  229796    226593   467348-   118551-    46083     93737-
23   08        257938-  445897    171460
24   01         33471-  359656    164752   403045-    51168-   215622-    92983
24   08         98267   444779    196759
25   01        126994-  461209    174691   476910-   155921-   247841    101789-
25   08         50131-  389918    162945-
```

```
FA 111   FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES

                         AFTER 12 ITERATIONS

ROW COL
 1   01     345340-   247786-   788684-     54953     69871-     57534-     23107
 1   08     300426     26857     87142-
 2   01     504353    776840-   171849      17216    103844-     39847-     41052
 2   08     113858-    28966     12798
 3   01     184507    869256    311862      60092-   136160      15689-      4404
 3   08      27132-     1262     27759-
 4   01      30063    805609    557822      91055     32843      12602      71233
 4   08     125097-     1702      2762-
 5   01     568100    266408-   717678      10779-    47043-      7350-     13265-
 5   08     277024-     2660       604-
 6   01     631055    696115    244326-     22555      5104       1280-     56048-
 6   08     202326      9567      9080-
 7   01     358408    426655-   775110-     82463-    27734-     22227-    118420-
 7   08     218512-     6801     12786
 8   01     129086    274763-   268510-    546210-   505910     180275     312069-
 8   08     196459-    28988-    31167-
 9   01     622734    691500    242429-     36993     16240      12821      10328-
 9   08     159089     26297-     6290
10   01     749429    351014-   426089-     52805-    84807-     47828-     60355-
10   08     252820-     5480     14202-
11   01       3428-    25606    999373-      1465       582-       569-      1637
11   08      15201-       21-      339
12   01     459983    429202-   601953-    173809-    64635      57354     135593-
12   08      61348-    15315-    53757-
```

TABLE 2.8.3

TABLE 2.8.4

```
FA 111   FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES

                               AFTER 12 ITERATIONS

ROW  COL
13   01       53914   280011-   896414-    63231-    72666    32161    48879-
13   08       75413    12287-    39662-
14   01      368744   739252-   329259-    67249     54922   133118    90386
14   08       11638-   26320     71385-
15   01      344162   757067-   410432-   170875    178577    14823-  232219
15   08       81933    15300      9771-
16   01      146169   465174-   463602-   265017    316617-  249471    77555
16   08       20437   122330-    74026-
17   01      113372   566441-   656111-   180700    241614-  198549    94805
17   08      108765    66707-    51069-
18   01      428273   369646-   543960-    40234-     9855-  147014    61110
18   08      132777    32973-   187574
19   01      533713   323337-   325155-    76436    186357    28185   183124
19   08      142490    62996-   157466
20   01      401720    82032-   501257-   129058    116738-   61062   114805
20   08      269303   108111-   215560
21   01      498677   181140    432255-   166147    181115-  215546    42465
21   08      211290    15606     13193-
22   01      136530   382861     93420-   624294-   151738-   39037   487084
22   08       95005   282694      7854
23   01      210870   237537     65617    471459-    59474-   68381-  388497
23   08       73827   591185-     2660
24   01      148873   359177     15373    397988-    43716    91227   427418
24   08       38633   109219-   226293
25   01      148067   471176     99710    470924-    53263-  111431-  336349
25   08       65940   122293-   260003-
```

FA 111 - FREQUENCY LIST OF DIFFERENCES BETWEEN P MATRICES

Amount of (±) Difference	Centroid versus 12th Iteration First factor	Centroid versus 12th Iteration Other factors
0.00	17	41
0.01	7	71
0.02	1	16
0.03		20
0.04		19
0.05		15
0.06		11
0.07		9
0.08		4
0.09		5
0.10+		14
(Max. diff.)		(0.174)

TABLE 2.9.1

TABLE 3.1.1

```
           FA 116     R     CORRELATION MATRIX       6 DECIMAL PLACES

      ROW  COL

        1  01     1000000      97600    594500-    713600     726700-    595400-    539700-
        1  08      705000-    496200-   383700-    605700-    629800-    595400-    539700-
        2  01       97600    1000000    696900-    734900     663600-    686100     649500-
        2  08      627400-    576700-   695800-    661700-    469600     686100     649500-
        3  01      594500-    696900-  1000000     860400-    931000     105300-    837100
        3  08      828400     787000    807800     865800      27400     105300-    837100
        4  01      713600     734900    860400-   1000000     940000-     58900     820000-
        4  08      911300-    748400-   732800-    871600-    109500-     58900     820000-
        5  01      726700-    663600-   931000     940000-   1000000      15000     859600
        5  08      929700     812000    770500     902300     153300      15000     859600
        6  01      595400-    686100    105300-     58900      15000    1000000     138700-
        6  08       23700      76500-   276800-     86400-    847600    1000000     138700-
        7  01      539700-    649500-   837100     820000-    859600     138700-   1000000
        7  08      812700     797900    778500     923700      12200-    138700-   1000000
        8  01      705000-    627400-   828400     911300-    929700      23700     812700
        8  08     1000000     759100    682900     884500     168200      23700     812700
        9  01      496200-    576700-   787000     748400-    812000      76500-    797900
        9  08      759100    1000000    765000     818700      30600      76500-    797900
       10  01      383700-    695800-   807800     732800-    770500     276800-    778500
       10  08      682900     765000   1000000     795000     149700-    276800-    778500
       11  01      605700-    661700-   865800     871600-    902300      86400-    923700
       11  08      884500     818700    795000    1000000      64900      86400-    923700
       12  01      629800-    469600     27400     109500-    153300     847600      12200-
       12  08      168200      30600    149700-     64900    1000000     847600      12200-

      DETERMINANT R EQUALS    .000000001409
```

TABLE 3.2.1

```
          FA 116 .                F MATRICES        6 DECIMAL PLACES

FACTR VAR  CENTROID           FACTOR VAR 40TH ITERATION    FACTOR VAR 59TH ITER.
   1  01      789129-            1  01       790685-          1  01       790246-
   1  02      593625-            1  02       589955-          1  02       589792-
   1  03      904690             1  03       906765           1  03       905742
   1  04      942851-            1  04       942083-          1  04       942302-
   1  05      978124             1  05       980417           1  05       981452
   1  06      101350             1  06       110769           1  06       110695
   1  07      873846             1  07       881319           1  07       881442
   1  08      928398             1  08       932825           1  08       933223
   1  09      821131             1  09       816964           1  09       816875
   1  10      763412             1  10       756932           1  10       756140
   1  11      931158             1  11       927726           1  11       927620
   1  12      251189             1  12       238599           1  12       238335
   2  01      533014             2  01       540712           2  01       539994
   2  02      744342-            2  02       741654-          2  02       740936-
   2  03      221396             2  03       224617           2  03       223984
   2  04      133906-            2  04       132236-          2  04       131665-
   2  05       95088             2  05        95570           2  05        95848
   2  06      932296-            2  06       972284-          2  06       973121-
   2  07      261095             2  07       256226           2  07       256099
   2  08       69929             2  08        71594           2  08        72076
   2  09      208917             2  09       196124           2  09       195298
   2  10      392102             2  10       382343           2  10       381700
   2  11      205367             2  11       206044           2  11       205890
   2  12      866667-            2  12       832106-          2  12       832406-
```

TABLE 3.2.2

```
            FA 116                 F MATRICES           6 DECIMAL PLACES

FACTR VAR  CENTROID          FACTOR VAR 40TH ITERATION    FACTOR VAR 59TH ITER.
    3  01     108488              3  01     131511             3  01     130662
    3  02     293332              3  02     280385             3  02     278942
    3  03      35326              3  03      46976             3  03      44077
    3  04     264813              3  04     294917             3  04     298354
    3  05      61214-             3  05      37410-            3  05      35515-
    3  06      74195              3  06     112429             3  06     112371
    3  07     173316              3  07     184254             3  07     183141
    3  08     137444-             3  08      81048-            3  08      78068-
    3  09     218824              3  09     168294             3  09     164212
    3  10     190948              3  10     110518             3  10     105851
    3  11     138448              3  11     128989             3  11     129943
    3  12      34360              3  12      12933             3  12      15786
    4  01      39523              4  01      14447-            4  01      13820-
    4  02      59883              4  02      31605             4  02      29295
    4  03     166784-             4  03     173094-            4  03     160974-
    4  04      16940              4  04      55337-            4  04      57838-
    4  05      39922-             4  05     130865-            4  05     138798-
    4  06       2093              4  06      29705-            4  06      27928-
    4  07     102584              4  07     176072             4  07     174857
    4  08     156319              4  08         98             4  08       8120-
    4  09      16187              4  09      11794             4  09      18434
    4  10     149193-             4  10       2264             4  10      15634
    4  11     157672              4  11     161837             4  11     163143
    4  12      37498              4  12       4063-            4  12       3275-
```

TABLE 3.3.1

FA 116 COMMUNALITIES 6 DECIMAL PLACES

VAR 1	2	3	4	5	6	7	ITERATION
943843	992934	950451	980238	972243	955269	898585	4
941887	988752	944717	983127	973579	960467	902597	8
940277	985429	937765	985540	975397	964236	903263	12
939066	983291	930686	987515	977517	966586	903839	16
938178	981885	924257	989155	979765	968003	904640	20
937455	980848	918776	990610	981946	968910	905530	24
936801	979954	914168	991928	983979	969604	906398	28
936207	979179	910358	993099	985818	970116	906948	32
935592	978410	907342	994140	987418	970669	907314	36
935056	977712	904844	995045	988876	971129	907326	40
934578	977103	902797	995791	990074	971542	907339	44
933808	976152	900091	996963	991865	972196	907241	52
933732	976044	899785	997075	992013	972274	907098	53
933656	975938	899577	997217	992241	972338	906929	54
933576	975837	899245	997309	992390	972405	906903	55
933504	975754	899059	997397	992555	972464	906815	56
933447	975672	898833	997464	992740	972516	906837	57
933402	975596	898583	997558	992882	972568	906735	58
933346	975508	898393	997629	992961	972625	906642	59
933280	975458	898270	997722	993058	972676	906722	60
933231	975405	898122	997769	993160	972721	906746	61
933185	975339	897998	997846	993224	972768	906685	62
933139	975290	897903	997933	993406	972794	906505	63
933090	975210	897663	998023	993538	972843	906456	64
933030	975146	897490	998098	993678	972885	906479	65

TABLE 3.3.2

FA 116 COMMUNALITIES 6 DECIMAL PLACES

VAR 8	9	10	11	12	ITERATION
885413	734794	746625	950823	739679	4
883808	736198	745818	948531	743899	8
882667	736687	744166	948735	746480	12
881918	736939	741970	948713	747789	16
881547	736896	739547	948230	748440	20
881428	736559	737256	947559	748788	24
881458	736045	735281	946880	749027	28
881578	735454	733655	946416	749180	32
881701	734900	732382	945998	749370	36
881857	734357	731352	945959	749514	40
881976	733913	730571	945873	749649	44
882125	733214	729517	946012	749830	52
882144	733151	729429	946188	749850	53
882171	733023	729293	946268	749861	54
882194	732955	729197	946323	749882	55
882217	732901	729112	946348	749903	56
882238	732823	729029	946278	749928	57
882254	732802	728944	946436	749937	58
882261	732731	728891	946370	749963	59
882255	732705	728847	946309	749968	60
882253	732686	728799	946427	749981	61
882272	732631	728749	946515	749990	62
882296	732588	728677	946527	749985	63
882305	732513	728596	946595	749995	64
882310	732438	728512	946596	750007	65

TABLE 3.4.1

```
FA 116    I - F PRIME  R INVERSE  F  MATRIX      6 DECIMAL PLACES
 ROW  COL                                                ITERATION
  1   01        8013      1649-     12070                    4
  2   01        1652-    17928      27965      3268          4
  3   01       12071     27965     125369      4128          4
  4   01        3268      4128       7327-     7327-         4
                                           284761
  1   01        7911      1319-     13319      2118          8
  2   01        1323-    18279      24169      2698          8
  3   01       13320     24169     128815      5935-         8
  4   01        2117      2698       5935-   294324          8

  1   01        7764      1025-     14344      1053         12
  2   01        1028-    18292      21490      1243         12
  3   01       14344     21490     128570      4432-        12
  4   01        1053      1243       4433-   301259         12

  1   01        7593       785-     15124       534         16
  2   01         789-    18147      19743       439         16
  3   01       15127     19742     126026      1535-        16
  4   01         533       438       1535-   303520         16

  1   01        7383       602-     15652       728         20
  2   01         605-    17987      18464       237         20
  3   01       15653     18464     122431      3286         20
  4   01         728       237       3287    300539         20

  1   01        7140       464-     16003      1424         24
  2   01         467-    17841      17353       451         24
  3   01       16004     17353     118100      9517         24
  4   01        1423       451       9517    292875         24

  1   01        6871       358-     16231      2397         28
  2   01         361-    17698      16300       888         28
  3   01       16231     16300     113387     16550         28
  4   01        2397       888      16551    282024         28

  1   01        6595       275-     16364      3527         32
  2   01         277-    17575      15326      1396         32
  3   01       16367     15326     108612     23757         32
  4   01        3526      1397      23757    269520         32
```

TABLE 3.4.2

```
FA 116    I = F PRIME  R INVERSE  F  MATRIX        6 DECIMAL PLACES
 ROW  COL                                                ITERATION
  1   01        6333      210-     16470      4726        36
  2   01         213-   17447      14405      1985        36
  3   01       16472    14404     104046     30941        36
  4   01        4728     1985      30941    256793        36

  1   01        6066      165-     16495      5932        40
  2   01         168-   17330      13555      2528        40
  3   01       16497    13554      99628     37565        40
  4   01        5933     2528      37565    243439        40

  1   01        5837      127-     16516      7030        44
  2   01         131-   17223      12841      3008        44
  3   01       16518    12842      95824     43298        44
  4   01        7031     3008      43298    232032        44

  1   01        5461       82-     16517      8742        52
  2   01          85-   17027      11657      3831        52
  3   01       16519    11656      89354     52228        52
  4   01        8741     3830      52228    212251        52

  1   01        5424       78-     16499      8839        53
  2   01          82-   17006      11533      3870        53
  3   01       16502    11532      88611     52893        53
  4   01        8838     3870      52894    210439        53

  1   01        5375       77-     16500      9109        54
  2   01          81-   16983      11383      4011        54
  3   01       16501    11382      87794     54166        54
  4   01        9109     4010      54166    207848        54

  1   01        5344       72-     16496      9239        55
  2   01          74-   16966      11286      4033        55
  3   01       16498    11285      87257     54773        55
  4   01        9239     4032      54773    206163        55

  1   01        5309       70-     16491      9434        56
  2   01          74-   16947      11191      4124        56
  3   01       16493    11191      86755     55603        56
  4   01        9435     4124      55603    204363        56
```

```
FA 116   I - F PRIME  R INVERSE  F  MATRIX        6 DECIMAL PLACES
 ROW  COL                                                ITERATION
  1   01        5276        70-      16475       9689        57
  2   01         73-      16937      11119       4208        57
  3   01       16476      11119      86359      56491        57
  4   01        9689       4208      56491     202546        57

  1   01        5237        70-      16445       9793        58
  2   01         73-      16920      11006       4239        58
  3   01       16446      11006      85658      57029        58
  4   01        9794       4240      57029     200666        58

  1   01        5223        60-      16485       9909        59
  2   01         64-      16907      10945       4284        59
  3   01       16487      10944      85400      57739        59
  4   01        9909       4284      57739     200196        59

  1   01        5202        58-      16503      10001        60
  2   01         61-      16888      10852       4359        60
  3   01       16505      10851      84883      58371        60
  4   01       10000       4359      58371     198895        60

  1   01        5170        62-      16450      10067        61
  2   01         66-      16871      10774       4372        61
  3   01       16453      10773      84395      58584        61
  4   01       10066       4373      58584     197479        61

  1   01        5154        58-      16462      10105        62
  2   01         62-      16855      10691       4391        62
  3   01       16464      10690      83926      58978        62
  4   01       10105       4392      58978     196523        62

  1   01        5117        61-      16454      10363        63
  2   01         65-      16841      10594       4533        63
  3   01       16456      10594      83410      59994        63
  4   01       10364       4533      59995     194524        63

  1   01        5084        58-      16446      10487        64
  2   01         61-      16824      10498       4585        64
  3   01       16447      10497      82835      60661        64
  4   01       10487       4584      60662     192915        64

  1   01        5056        56-      16442      10660        65
  2   01         61-      16812      10422       4660        65
  3   01       16443      10421      82381      61345        65
  4   01       10660       4660      61345     191231        65
```

TABLE 3.4.3

```
FA 116    LOG DET I  -  F PRIME  R INVERSE  F        8 DECIMAL PLACES

ITERATION  1 FACTOR            2 FACTORS           3 FACTORS           4 FACTORS

     4     209620485-          385098839-          511478317-          567646395-
     8     210176861-          384507490-          501365948-          555071222-
    12     210991447-          385089429-          499383934-          551659505-
    16     211958659-          386274862-          501288715-          553097894-
    20     213176713-          387801731-          504856558-          557081939-
    24     214630178-          389563785-          509465972-          562920696-
    28     216298005-          391552235-          514898309-          570268052-
    32     218078519-          393618648-          520850698-          578684547-
    36     219839051-          395684402-          527204352-          587906357-
    40     221709759-          397842434-          534038458-          598115069-
    44     223381030-          399777715-          540702445-          608133025-
    52     226272782-          403164069-          553927148-          628331429-
    53     226568031-          403512921-          555404505-          630544655-
    54     226962153-          403965826-          557440928-          633713705-
    55     227213355-          404257968-          558729312-          635704095-
    56     227498727-          404592006-          560109485-          637887797-
    57     227769521-          404888436-          561235597-          639825688-
    58     228091742-          405254272-          562815116-          642225961-
    59     228207997-          405403910-          563846483-          643728900-
    60     228382965-          405627714-          565286489-          645903096-
    61     228650945-          405939436-          566286374-          647457944-
    62     228785558-          406115258-          567508187-          649257876-
    63     229098458-          406464249-          569196978-          652036032-
    64     229379445-          406789100-          570906181-          654685896-
```

TABLE 3.5.1

TABLE 3.5.2

```
FA 116                SUM LOG ONE MINUS H SQUARE              8 DECIMAL PLACES

ITERATION  1 FACTOR           2 FACTORS          3 FACTORS          4 FACTORS

     4     691884729-         1115410271-        1343369156-        1431637249-
     8     692512050-         1118429213-        1335903510-        1419559009-
    12     693306605-         1120872477-        1334647129-        1416419350-
    16     694251436-         1122782262-        1336384565-        1418026409-
    20     695260255-         1124329901-        1339162766-        1422123898-
    24     696245993-         1125637792-        1342444805-        1428056018-
    28     697170106-         1126820176-        1346010546-        1435442297-
    32     698015244-         1127859496-        1349684325-        1443931991-
    36     698756198-         1128851327-        1353403320-        1453218981-
    40     699454400-         1129782707-        1357132728-        1463451799-
    44     700020410-         1130565814-        1360578092-        1473488965-
    52     700905270-         1131823141-        1366869080-        1493723094-
    53     700996188-         1131961864-        1367544810-        1495939109-
    54     701121966-         1132138313-        1368437199-        1499082366-
    55     701195253-         1132242284-        1368992260-        1501127115-
    56     701284198-         1132371781-        1369552771-        1503308167-
    57     701368258-         1132490897-        1370008720-        1505269809-
    58     701455067-         1132617434-        1370675579-        1507659342-
    59     701489085-         1132671846-        1371068405-        1509139549-
    60     701543652-         1132761413-        1371684118-        1511314031-
    61     701608262-         1132863697-        1372083601-        1512876028-
    62     701657234-         1132936473-        1372599853-        1514691443-
    63     701763875-         1133077571-        1373251498-        1517474479-
    64     701829959-         1133171634-        1373873292-        1520092832-
```

```
                 FA 116   CONVERGENCE TEST

              LOG DET
  ITER-     I - F PRIME      SUM LOG                        CHANGE FROM
  ATION     R INVERSE F      1 - H SQ       DIFFERENCE        PREVIOUS
                                                             ITERATION
     4      - 5.67646       - 14.31637       8.63991
     8      - 5.55071       - 14.19559       8.64488          0.00497
    12      - 5.51659       - 14.16419       8.64760          0.00272
    16      - 5.53097       - 14.18026       8.64929          0.00169
    20      - 5.57082       - 14.22124       8.65042          0.00113
    24      - 5.62921       - 14.28056       8.65135          0.00093
    28      - 5.70268       - 14.35442       8.65174          0.00039
    32      - 5.78685       - 14.43932       8.65247          0.00073
    36      - 5.87906       - 14.53209       8.65303          0.00056
    40      - 5.98115       - 14.63452       8.65337          0.00034
    44      - 6.08133       - 14.73489       8.65356          0.00019
    52      - 6.28331       - 14.93723       8.65392          0.00036
    53      - 6.30545       - 14.95939       8.65394          0.00002
    54      - 6.33713       - 14.99082       8.65369          0.00025-
    55      - 6.35704       - 15.01127       8.65423          0.00054
    56      - 6.37888       - 15.03308       8.65420          0.00003-
    57      - 6.39826       - 15.05270       8.65444          0.00024
    58      - 6.42226       - 15.07660       8.65433          0.00011-
    59      - 6.43729       - 15.09139       8.65410          0.00023-
    60      - 6.45903       - 15.11314       8.65411          0.00001
    61      - 6.47468       - 15.12876       8.65408          0.00003-
    62      - 6.49258       - 15.14691       8.65433          0.00025
    63      - 6.52036       - 15.17474       8.65438          0.00005
    64      - 6.54686       - 15.20093       8.65407          0.00031-
  DETERMINANT R EQUALS 0.000000001409
```

TABLE 3.5.3

FA 116 - DETERMINATION OF THE NUMBER OF FACTORS

$N = 138$, $p = 12$, $|R| = 1.409 \times 10^{-9}$

Degrees of freedom $= p(p-1)/2 = 66$, $\chi^2_{.95}(66) = 78$ (approx.)

$\log_{10}|R| = 0.14891-9 = -8.85109$

$m_3 = N - k - (2p+11)/6 = 129.16667$, $m_4 = 128.16667$

	Iteration 4		Iteration 59	
	3 factors	4 factors	3 factors	4 factors
$\log_{10} \lvert I - F'R^{-1}F \rvert$	- 5.11478	- 5.67646	- 5.63846	- 6.43729
$\Sigma \log_{10}(1-h^2)$	-13.43369	-14.31637	-13.71068	-15.09139
Difference	8.31891	8.63991	8.07222	8.65410
$\log_{10}\lvert R \rvert$	- 8.85109	- 8.85109	- 8.85109	- 8.85109
$\log_{10}\lvert R(\underline{z}\vert\underline{x})\rvert$	- 0.53218	- 0.21118	- 0.77887	- 0.19699
$\log_e\lvert R(\underline{z}\vert\underline{x})\rvert$	- 1.22539	- 0.48626	- 1.79341	- 0.45359
$-m_k\log_e\lvert R(\underline{z}\vert\underline{x})\rvert$	158.27954	62.32232	231.64878	58.13512
$\sqrt{2\chi^2} \quad -\sqrt{2\nu-1}$	6.35	- 0.28	10.08	- 0.66
Probability that k factors sufficient	< 0.00001	0.610	< 0.00001	0.745

TABLE 3.5.4

```
FA 116   D SUB LAMBDA - - CHARACTERISTIC ROOTS OF  F PRIME F      6 DEC PLACES

                            FOR CENTROID SOLUTION                 ROW  COL

 7704695                                                           1   01
          2612417                                                  2   01
                   328295                                          3   01
                            114563                                 4   01

                            FOR FORTIETH ITERATION                ROW  COL

 7696157                                                           1   01
          2629715                                                  2   01
                   297827                                          3   01
                             99305                                 4   01

                          FOR FIFTY-NINTH ITERATION               ROW  COL

 7695002                                                           1   01
          2629967                                                  2   01
                   296384                                          3   01
                             95945                                 4   01
```

TABLE 3.6.1

TABLE 3.7.1

```
FA 116   P MATRICES  - -  PRINCIPAL AXES SOLUTIONS       6 DECIMAL PLACES
ROW  COL             AFTER CENTROID

  1   01       654933-    685379    135804     55099
  2   01       744172-   610437-    263041     21195
  3   01       930873      22926     33562   168427-
  4   01       951688-     63222    259998      2248
  5   01       975847    116558-    60321-    40099-
  6   01       104937-   933729-     31560    33160-
  7   01       908150      52926    190336     92596
  8   01       922477    133957-   124746-    160214
  9   01       844943      14708    227811      2170
 10   01       828815     212830    197754   154674-
 11   01       952236     14342-    156533    147637
 12   01        55945    902010-     3041-      5735
ROW  COL           AFTER FORTIETH ITERATION

  1   01       658717-    690349    149615    46702-
  2   01       739509-   610723-    240392     8207-
  3   01       931063      20341     23024   192411-
  4   01       953965-     57838    265354   106009-
  5   01       977221    116638-    56373-   130895-
  6   01       102366-   978055-     54237    33410-
  7   01       914227      52317    234028    118363
  8   01       927701    126490-    71930-      7323
  9   01       837086       7054    179585    36661-
 10   01       819248     204475    130423    36934-
 11   01       949595      3851-    174688    117038
 12   01        54470    863475-    29623-      8885
ROW  COL         AFTER FIFTY-NINTH ITERATION

  1   01       658731-    689217    148587    48177-
  2   01       739208-   610432-    237330    11193-
  3   01       929675      19994     20262   182445-
  4   01       954526-     57545    266012   111518-
  5   01       977907    116769-    58202-   140129-
  6   01       102502-   978979-     52875    30382-
  7   01       914471      52355    235023    111395
  8   01       928111    126069-    70527-     2069-
  9   01       836806       6435    176870    34104-
 10   01       818399     204333    128924    27233-
 11   01       949640      3805-    178216    113048
 12   01        54311    863792-    27462-     11053
```

TABLE 3.8.1

```
FA 116   FE MATRIX   - - -   LAWLEYS SOLUTION -- 6 DECIMAL PLACES
                                  AFTER CENTROID
ROW COL
  1  01        296872-   887109    208964     37388
  2  01        974836-   212559-    24091       506
  3  01        814949    423615-   125697    192757-
  4  01        856573-   472176    142083     21370-
  5  01        811198    556356-    32671     49677-
  6  01        510751-   764651-   197028-    24066-
  7  01        771359    414231-   322986     37931
  8  01        768623    541966-     5272-   160009
  9  01        692413    420356-   326834     55621-
 10  01        774394    232828-   313784    207141-
 11  01        786637    490309-   290699     97740
 12  01        349687-   809830-   196030-    16126

                                  AFTER 40 ITERATIONS
ROW COL
  1  01        684998-   708806     86969        81-
  2  01        761601-   558303-    47028     57679
  3  01        905884     26981    265875    115259-
  4  01        989125-   108102    121813     22176-
  5  01        968896    123429-   167333     95060-
  6  01         98174-   952880-    21320-    20557-
  7  01        862509     84888    310089    256177
  8  01        927491    139825-    82561     18786
  9  01        791548     36009    315972     89146
 10  01        779855    222193    283479     68006
 11  01        908299     19425    260977    229571
 12  01         70561    856005-    79280-    11481-

                                  AFTER 59 ITERATIONS
ROW COL
  1  01        705388-   656417      3772     69823-
  2  01        744492-   645737-    64913      7079-
  3  01        895937     37051     58079-   301569-
  4  01        994017-     8259      8387-    97062-
  5  01        966445     90732-    68407-   214557-
  6  01         70435-   983125-    12268-    31288
  7  01        851253     64191    303763    292606-
  8  01        927118     95428-    30621    112560-
  9  01        781802     15193    145932    316214-
 10  01        765151    215567    123687    285776-
 11  01        899341     13327    273588    250057-
 12  01         95884    856819-    12742-    80420
```

FA 116 - FREQUENCY LIST OF DIFFERENCES AMONG P MATRICES

Amount of (±) Difference	Centroid vs 40th Iter.		Centroid vs 59th Iter.		40th vs 59th Iter.	
	Factor 1	Others	Factor 1	Others	Factor 1	Others
0.00	8	8	8	7	12	30
0.01	4	10	4	12		6
0.02		3		4		
0.03		4		3		
0.04		4		3		
0.05		2		2		
0.06		0		0		
0.07		1		1		
0.08		0		0		
0.09		1		0		
0.10		3		4		
(Max. diff.)		(0.108)		(0.158)		

TABLE 3.9.1

APPENDIX C

IBM 650 PROGRAMS
FOR FACTOR ANALYSIS

FACTOR ANALYSIS PROGRAMS

by Rolf E. Bargmann and Robert H. Brown
Virginia Polytechnic Institute

Library Classification

Preliminaries:

6.6.020.1 Translate original data cards into a complete R matrix ready for input in (5.2.001.2) and (6.6.021.5).

5.2.001.2 (G 864) Obtain $|R|$ and R^{-1}

Factor Analysis Programs:

6.6.021.1 (A M 557) Centroid Solution, initial

6.6.021.2 Translate original data card into input for (A M 561)

6.6.021.3 (A M 561) Iteration of Centroid Solution

6.6.021.4 parts I and II
Convergence Tests and Decision on Number of Factors

6.6.021.5 Maximum-Likelihood Solution. Main Routine

6.6.021.6 (FAROT)
Rotation into Simple Structure

Special Routines (use only if Lawley's particular maximum-likelihood solution is desired or Principal Axes)

6.6.021.7 Obtain and punch $F'D_{1/1-h^2}F$ or $F'F$

5.2.005.2 Obtain characteristic roots and eigenvectors (subroutine only).

6.6.021.8 Punch results of 5.2.005.2.

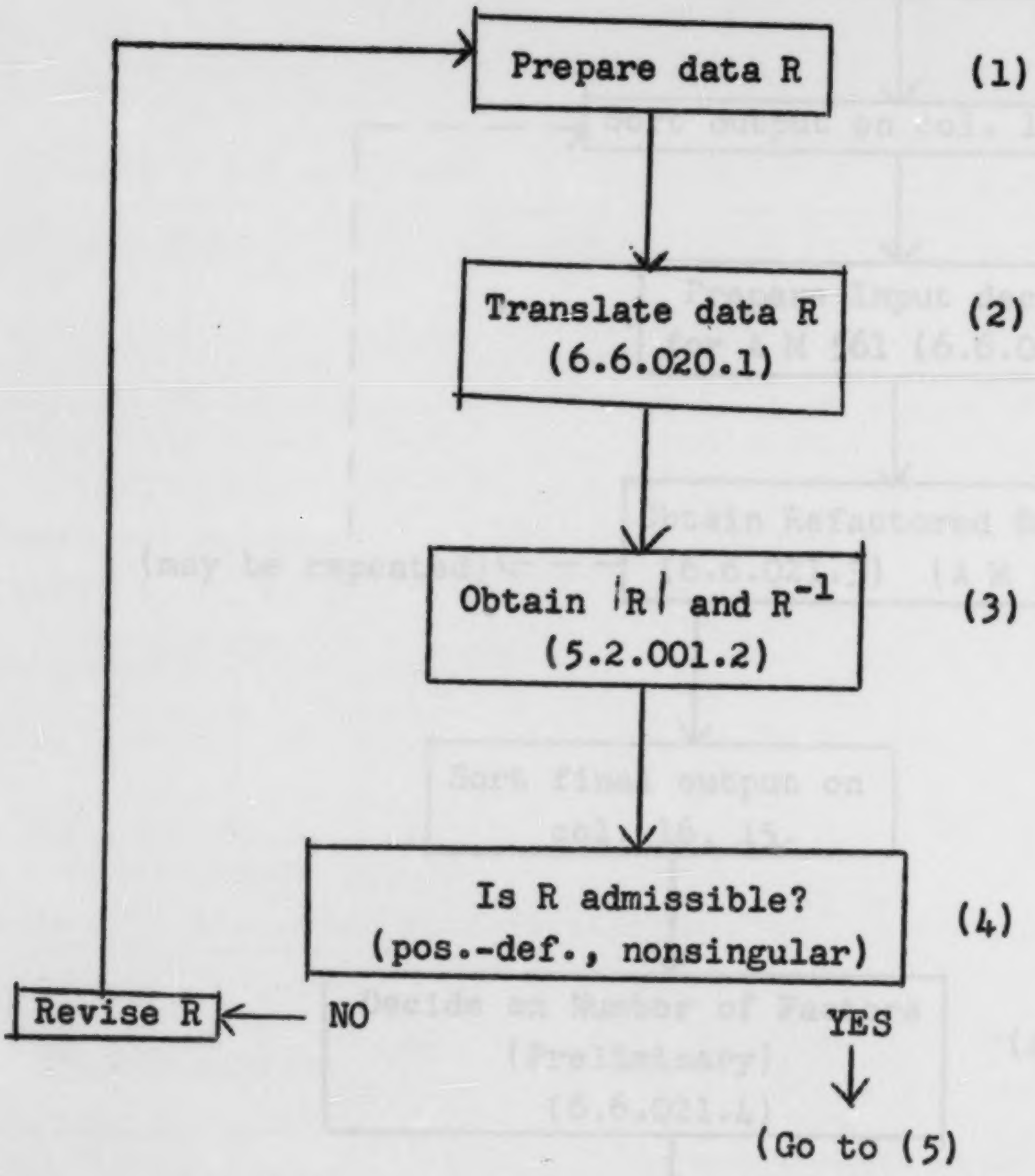
6.6.021.9 Obtain Lawley's Solution, or Principal Axes

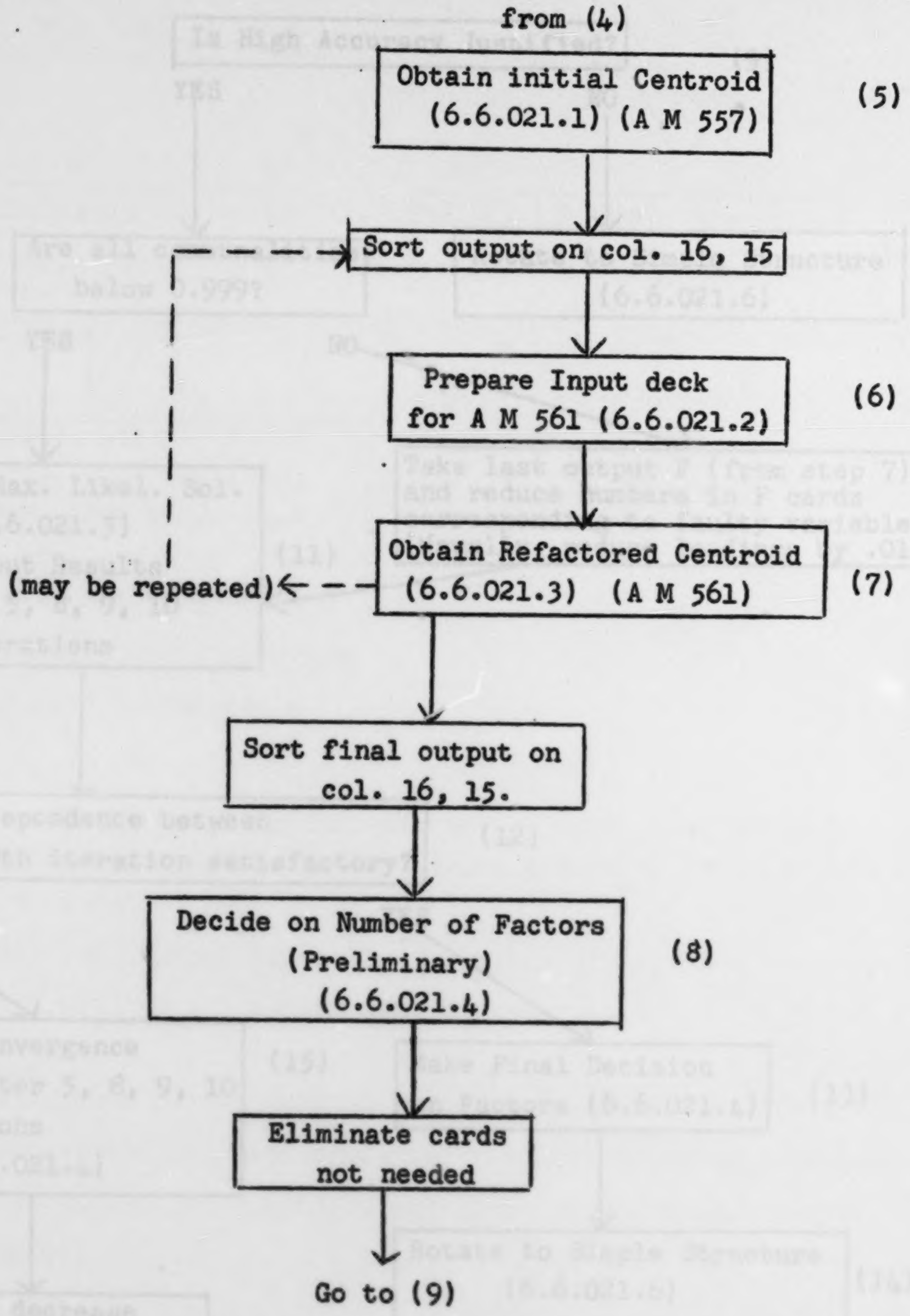
Limitations: 35 variables
10 factors

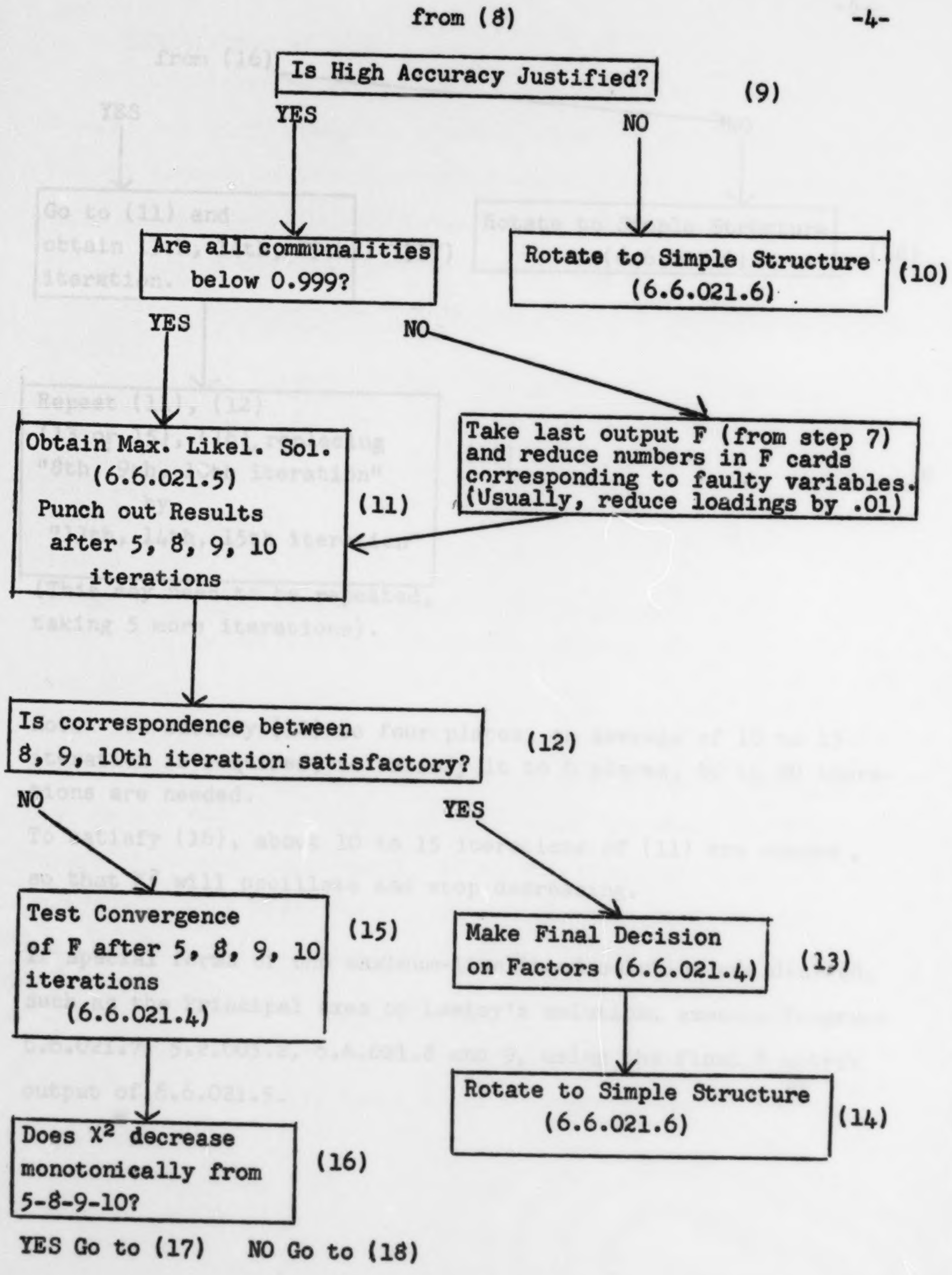
Factor Analysis Routines

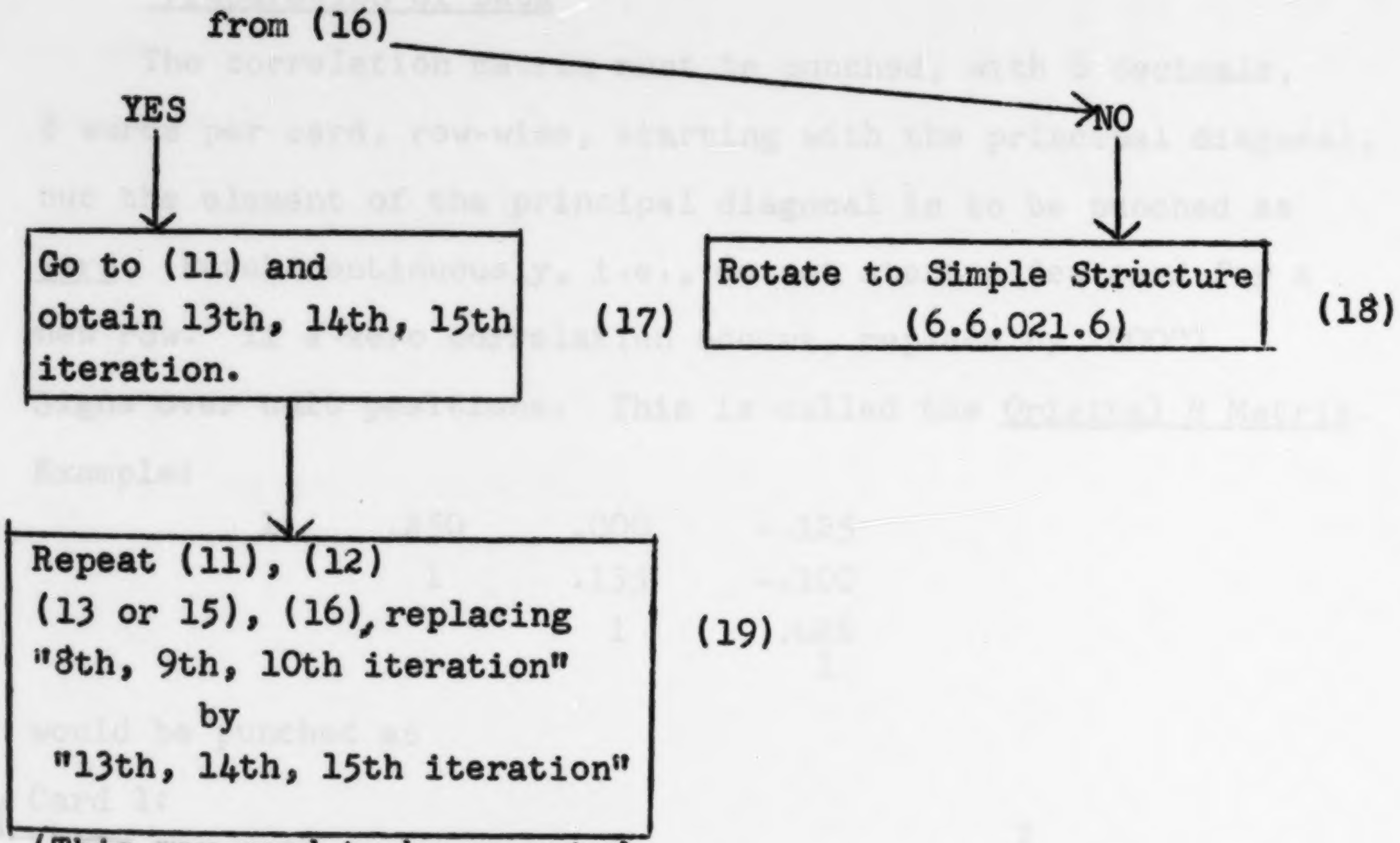
General Flow of Programs

Limitations 35 variables 10 factors.




from (4)
Obtain initial Centroid (6.6.021.1) (A M 557)
(5)
Sort output on col. 16, 15
Prepare Input deck for A M 561 (6.6.021.2)
(6)
Obtain Refactored Centroid (6.6.021.3) (A M 561)
(7)
(may be repeated)
Sort final output on col. 16, 15.
Decide on Number of Factors (Preliminary) (6.6.021.4)
(8)
Eliminate cards not needed
Go to (9)


from (8)
Is High Accuracy Justified?
(9)
YES
NO
Are all communalities below 0.999?
Rotate to Simple Structure (6.6.021.6)
(10)
YES
NO
Obtain Max. Likel. Sol. (6.6.021.5)
Punch out Results after 5, 8, 9, 10 iterations
(11)
Take last output F (from step 7) and reduce numbers in F cards corresponding to faulty variables. (Usually, reduce loadings by .01)
Is correspondence between 8, 9, 10th iteration satisfactory?
(12)
NO
YES
Test Convergence of F after 5, 8, 9, 10 iterations (6.6.021.4)
(15)
Make Final Decision on Factors (6.6.021.4)
(13)
Does X^2 decrease monotonically from 5-8-9-10?
(16)
Rotate to Simple Structure (6.6.021.6)
(14)
YES Go to (17)
NO Go to (18)


from (16)
YES
NO
Go to (11) and obtain 13th, 14th, 15th iteration.
(17)
Rotate to Simple Structure (6.6.021.6)
(18)
Repeat (11), (12) (13 or 15), (16), replacing "8th, 9th, 10th iteration" by "13th, 14th, 15th iteration"
(19)

(This may need to be repeated, taking 5 more iterations).

Note: To satisfy (12) to four places, an average of 10 to 15 iteration is required; to satisfy it to 6 places, 60 to 80 iterations are needed.

To satisfy (16), about 10 to 15 iterations of (11) are needed, so that γ^2 will oscillate and stop decreasing.

If special forms of the maximum-likelihood solution are desired, such as the Principal Axes or Lawley's solution, execute Programs 6.6.021.7, 5.2.005.2, 6.6.021.8 and 9, using the final F matrix output of 6.6.021.5.

Preparation of Data

The correlation matrix must be punched, with 6 decimals, 8 words per card, row-wise, starting with the principal diagonal; but the element of the principal diagonal is to be punched as zero. Punch continuously, i.e., do not start a new card for a new row. If a zero correlation occurs, replace by .00001 Signs over unit positions. This is called the Original R Matrix.

Example:

$$\begin{array}{cccc}
1 & .250 & .000 & -.125 \\
 & 1 & .135 & -.100 \\
 & & 1 & .026 \\
 & & & 1
\end{array}$$

would be punched as

Card 1:

```
                                               X
0000000000;  0000250000;  0000000001;  0000125000;
                                   X
0000000000;  0000135000;  0000100000;  0000000000
```

Card 2:

```
0000026000;  0000000000;  zeros.
```

Even if the last card contains only the last diagonal entry it must be punched, i.e., the last card would then have all zeros. This will happen, for example, if p (number of variables) is 14, because, in that case, 105 elements would have to be punched, and the 105'th (last principal diagonal) would start a new card.

The order of cards is important and care must be taken that it is not destroyed. (This form of input, though inconvenient to us, is required for the Texas A and M programs, and we did not change it).

Except for Header Cards, these are all the cards that need to be prepared.

We will always designate

n = sample size or degrees of freedom due to error plus 1 if the correlation matrix was obtained from a design.

p = number of variables.

k = number of factors.

6.6.020.1 Obtain complete R matrix

a. Input: original R matrix, fixed decimal (6 places), 8 words per card, punched row-wise (one triangle of symmetric matrix)

b. Program: 6.6.020.1

c. 533 Control Panel: 80 - 80 Board

d. Header card: data card (no overpunches necessary)

Col's	Field	Form
1-10	p =#of variables(rows in R)	00 00(p) 0000
11-20	blank	
21-30	problem identification	00 0000 OIII
31-70	Zeros	
71-80	card identification	

e. Console set-up: (1) Storage Entry switches to 70 1951 1900
(2) Standard settings otherwise*

f. Input deck: (1) Program deck
(2) Header card
(3) Original R matrix cards

g. Stops: (1) 01 0001 1950 End
(2) 01 0002 8000 Header card not entered

*Standard settings are, for this write-up, considered to be:
Programmed Switch - stop, Display Switch - Program Register
Half-cycle Switch - Run;Overflow Switch - Stop
Control Switch - Run;Error Sense Sw. - Stop

h. Output: a complete R matrix, in proper form for input to the North Carolina State matrix inversion program. (5.2.001.2)

Col's	Field
1-2	#of decimal places (6)
3-4	row#of the elements in the card
5-6	column#of the first element in the card
7	#of elements in the card
8-10	problem identification (from the Header card)
11-20	first element
21-30	second element (sign is over unit's digit)
:	: : (for each element: 10,20...)
71-80	seventh element

Approx. time required: 5 minutes maximum

5.2.001.2 Obtain R^{-1} and $|R|$.

a. Input: Complete R matrix from 6.6.020.1

b. Program: Library Routine 5.2.001.2(North Carolina State Matrix Inversion Program).

c. 533 Control Panel: 80 - 80 Board

d. Load card: Program card, with "12 pch" in col. 1,10,20,..,80.

Col's	Field	Form
1-10	# of rows in R	00 0000 00 (p)
11-20	# of rows in R	00 0000 00 (p)
21-74	zeros	
75-80	problem identification	

e. Console set-up: (1) Storage Entry switches to 70 1952 2000

(2) Standard settings otherwise

f. Input deck:
(1) Program deck
(2) Control card
(3) R matrix deck from 6.6.020.1

NOTE: There are two parts to the program deck. Remove the last card of the first part and add the second part (to obtain determinant R). Has been done in assembled deck.

g. Stops: There are no programmed stops. A divide overflow usually indicates a singular matrix.

h. Output: $|R|$ and R^{-1}, floating decimal, first card has $|R|$ in word 1. Other cards:

Col's	Field
1-10	first element (1st two digits are exponent plus 50).
11-20	second element (sign over units position for all elements)
:	:
61-70	seventh element
71-72	row #
73-74	part#(7 elements per part, max. of 5 parts)
75-80	problem identification

Approx. time required: .00008 p^3 hours, plus loading and punching.

The output $|R|$ and R^{-1} should be tabulated so that at least a partial check can be made for singularity or near-singularity (due to the fact that some variables constitute mere mathematical combinations of others). Also positive-definiteness can be partially checked (lack of positive definiteness is almost always due to errors in computing R from data, or can be present if R is not obtained from experimental data but by some mathematical process).

a. If the characteristic of $|R|$ is less than 25, near-singularity is indicated and the culprit variables (mathematical constructs usually) should be eliminated.

b. If $|R|$ is negative or if the characteristic in a diagonal element of R^{-1} is 49 or less, or if a diagonal element is negative, R is not positive-definite and corrections must be made.

c. If the characteristic of a diagonal element of R^{-1} is greater than 54, near-singularity is indicated.

MAIN PROGRAMS

6.6.021.1 (A M 557) Obtain initial Centroid Solution.

a. Purpose: to obtain the initial estimate of factor loadings, i.e., an initial F matrix.

b. Program: Library Routine 6.6.021.1
(Texas A & M, A M 557)

c. 533 Control Panel: 80 - 80 Board

d. Header card: data card with p = number of variables (rows in R) in card col's 5-6 and zeros elsewhere

e. Console set-up:
  (1) Storage Entry to 70 1951 2000
  (2) Standard settings otherwise

f. Input:
  (1) Program cards (Lib. R. 6.6.021.1 (A M 557)).
  (2) Header card
  (3) Original R matrix cards (see Preparation of Data)

g. Stops: None

h. Output: first F matrix (F_A),

Col's	Field	Form
1-10	matrix element, sign in col. 10	0000.XXXXXX
11-20	row number	00 00rr 0000
21-30	column number	00 00cc 0000

NOTE: the output cards should be sorted by row number, card col's 16-15. Those having row number 00 should be discarded since they only contain the percent of factoring. This is the criterion used in the centroid program to determine when the factoring is complete. For more program detail, obtain original program writeup from Texas A and M College (we have none available).

Approx. time required: 1 hour for the largest problem;about 20 minutes for a medium-sized problem, say 25 variables and 5 factors.

6.6.021.2 Obtain new R matrix

a. Purpose: to obtain a new R matrix, having diagonal elements equal to the communalities of the initial F matrix, for use in obtaining an improved centroid solution.

b. Program: 6.6.021.2

c. 533 Control Panel: 80 - 80 Board

d. Header card: same one used in 6.6.021.1

e. Console set-up
   (1) Storage Entry to 70 1951 2000
   (2) Standard settings otherwise

f. Input:
   (1) Program cards 6.6.021.2
   (2) Transfer card (TR to 0100)
   (3) Header card:
   (4) F matrix cards from 6.6.021.1 sorted by row
   (5) Transfer card(TR to 0200)
   (6) Original R matrix (see section on Preparation of Data)

g. Stops: None

h. Output: new R matrix for input to 6.6.021.3 having the same format as the original R matrix.

NOTE: the new R matrix should be listed and the diagonal elements examined to see if they exceed 1.000000. See General Flow of Programs if they do.

Approx. time required: 5 mins maximum.

6.6.021.3 Improved centroid solution

a. Purpose: to obtain an improved F matrix for use as a starting point in the iterative solution.

b. Program: Library Routine 6.6.021.3 (A M 561).

c, 533 Control Panel: 80 - 80 Board

d. Header card: same one used in 6.6.021.1 and 2.

e. Console set-up: (1) Storage Entry to 70
1951 2000
(2) Standard settings otherwise

f. Input: (1) Program cards (Lib. R. 6.6.021.3 (A M 561).
(2) Header card
(3) New R matrix cards from 6.6.021.2

g. Stops: None

h. Output: new F matrix (F_B) for input to 6.6.021.5, having the same format as the output of 6.6.021.1

Approx. time required: 80 mins maximum; about 30 mins for 25 variables, and 5 factors

6.6.021.4 Convergence Test and Decision on Number of Factors. (2 parts).

a. Description of Program.

This program develops a X^2 goodness-of-fit test statistic, which is a slight modification of Bartlett's test. For exact probabilities, an F obtained by the maximum-likelihood program (6.6.021.5) is required; the latter, however, requires a preliminary decision on the appropriate number of factors. Such a preliminary decision can be made, conveniently, on the basis of this program (6.6.021.4) using the F matrix obtained in 6.6.021.3 in lieu of a maximum-likelihood F. In more than twenty studies made to-date, not a single instance occurred where the preliminary decision had to be modified after the final F (max. likelihood) was known. In other words, the once refactored centroid is so close to the maximum-likelihood solution that the probabilities of this test do not change appreciably. For a given F (correlations between observed variables $\underline{z}$, and artificial variables (factors) $\underline{x}$) the program computes $|R(\underline{z}|\underline{x})|$, i.e., the determinant of the partial correlation matrix of $\underline{z}$ given $\underline{x}$. It then develops the X^2 - statistic which tests partial independence, and thus evaluates the probability that the observed value of the determinant or a larger one could have arisen by chance if a model of 0 factors (independence), 1 factor, ... up to k factors (maximum k = 10) were true. The output contains the entire sequence of X^2's for 0, 1, 2 . . . k factors, along with the degrees of freedom and, since the latter are usually very large ($p(p-1)/2$), it also lists the normal curve equivalents obtained from the square-root approximation of X^2 for large degrees of freedom. Thus if these normal curve equivalents are -2 or smaller for a stated number of factors, that number of factors is not sufficient. If it is around zero, the stated number of factors has appreciable plausibility, and should be adopted. In the preliminary decision these normal curve equivalents must be smaller for a given number of factors than in the final one for the same number of factors. Thus, even if a normal curve equivalent should be -1 for 4 factors, say and +1 for 5 factors, four factors should be accepted as a plausible model. Usually (with no exception in our practice to date) the normal curve equivalent turn from values below -4 for (x-1) factors to around zero for x factors, to above +3 for (x+1) factors, even in the preliminary decision (F from 6.6.021.3) so that there can be no question regarding the proper number of factors. In practice, in the preliminary decision, we took k* as large as the number of factors punched out in 6.6.021.3, which has been, in all instances, extreme overfactoring. If 6.6.021.3 should punch out more than 10 factors, the cards for the 11th factor and above must be discarded, because this and the following programs are limited to 10 factors because of storage restrictions in the IBM 650.

---

* for Header Card, i.e., 6.6.021.4 will punch out results up to k factors which is usually way beyond the turning point.

b. Mathematical procedure

The test statistic is

$$\chi^2 = -m_k \ln |R(\underline{z}|\underline{x})|$$

with $p(p-1)/2$ d.f., where

$m_k = n - (2p + 11)/6 - k$, and where

n is the sample size and must be replaced by (degrees of freedom due to error plus one) if the correlation matrix was obtained from the error sum-of-squares and products in a design. The statistic can be reduced to the form

$$\chi^2 = -m_k [\ell n|R| + \ell n|I-F'R^{-1}F| - \sum_{i=1}^{p} \ell n(1-h_i^2)] .$$

Part I of the program stores $|R|$ in floating form, and develops the matrix $I-F'R^{-1}F$, the latter operation taking the longest time of the entire program. In part II, one obtains the log-determinants of the parts of the $I-F'R^{-1}F$ matrix corresponding to 1, 2, . . . k factors, and the h_i^2's corresponding to 1, 2, . . . k factors. They are combined into the χ^2 - statistic for 1, 2, . . . k factors. The number of factors, χ^2, degrees of freedom, and normal curve equivalents are listed in the output for zero factors (test of independence), 1, 2, . . . k factors, one card for each.

c. Program: Library Routine 6.6.021.4 parts I and II

d. Control Panel: 80 - 80 board

e. Console set-up: (1) Storage entry switches
to 70 1951 2000
(2) Overflow Switch to Sense
(3) Standard Settings otherwise

f. Header Card (load card)

12-punches in 1, 10, 20, 30, 40, 50

Word 1: 00 0000 1700
Word 2: 00 [nnnn] 0000
Word 3: 00 [00pp] 0000
Word 4: 00 [00kk] 0000
Word 5: any 10 digit numerical identification,

where nnnn = sample size, or degrees of freedom due to error plus one. Maximum number 9999.
pp = number of variables, maximum 35
kk = number of factors for which output is desired, maximum 10.

g. Input:
(1) Program deck part I
(2) Header card (load card)
(3) Cards for $|R|$ and R^{-1} in floating form, i.e., output of 5.2.001.2, making sure that $|R|$ is the first card, R^{-1} follows (in arbitrary order).
(4) Transfer card to 1800
(5) F matrix, in any order, output from 6.6.021.1, 6.6.021.3, or 6.6.021.5; for the latter, however, make sure that the five communality cards (end of output) are discarded.
(6) Transfer card to 1850
(7) Program deck part II

Press End-of-File key when the last card of Program deck part II is to be read, transfer to execution is automatic. A considerable time (around 10 minutes) may pass before Part II is read, during which Part I is executed.

h. Stops: 01 0123 in display lights, means that a term of R^{-1} exceeds 10^5. The value, in floating form, can be displayed from upper accumulator. Check R^{-1} matrix, R is probably near singular (see General Flow of Programs).

01 0666 in display lights, means that $|R|$ is negative. Probably, $|R|$ was not the first card in data input (step 3, above). Check order.

01 0802 in display lights means that for a certain variable, at a certain stage, h^2 was greater than 1. The value of h^2 is in upper accumulator (6 decimals). The variable number is in data address of the lower accumulator. The distributor contains, in data address, the largest number of factors for which h^2 was less than unity for all variables. Depressing Program Start key will produce output for all factors up to the last legitimate stage.

01 1820 in display lights means that the square root of a negative number is to be extracted. It would occur if $1 - h^2$ were negative, but this condition would have been checked in the previous stop. This stop would also occur if $p(p-1) -1$ were less than zero, which probably indicates a mistake in the Header card. Depressing Program Start switch will bypass this root and thus probably produce an erroneous result.

i. Output:

Word 1: Identification from Header Card

Word 2: Number of factors; 0, 1, 2, . . . k

Word 3: χ^2 with 4 decimals

Word 4: degrees of freedom (data address)

Word 5: Normal curve equivalents, with 4 decimals

Words 6-8: zeros.

NOTE: If p is very small (less than 7, say) it may be preferable to read probabilities from a table of the distribution of χ^2 with degrees of freedom stated in word 4 (e.g., from Table 7 of the Biometrika Tables for Statisticians). Read upper tail probabilities. For p of the usual size, the probabilities corresponding to the normal curve equivalents in word 5 are quite satisfactory. Notice that these probabilities are exact only for k factors after F has been obtained from the maximum-likelihood

after F has been obtained from the maximum-likelihood

program (6.6.021.5) based upon the preliminary decision assuming k factors. If k has been adopted in a preliminary decision, based on the (6.6.021.3) output (and this has always been correct in our past experience) the present program may be used with an F matrix obtained after, say, 5, 8, 9, 10 iterations of the maximum-likelihood program. If the χ^2 - value in word 3 of output of that card which has k in word 2 is steadily decreasing, even by small amounts, from iteration 8 to

9 to 10, this will indicate that convergence has not been reached and that more iterations of program (6.6.021.5) are needed. See General Flow of Programs.

Average time: 11 minutes for $p = 25$ $k = 10$, maximum size 20 minutes.

6.6.021.5 Maximum-Likelihood F.

a. Purpose: to obtain a maximum-likelihood estimate of the factor matrix F, by iteration.

b. Program: 6.6.021.5

c. 533 Control Panel: 80 - 80 board

d. Header Card (Data card - no over-punches)

Word 1: 00 00pp 0000
Word 2: 00 00kk 0000
Word 3: 00 0000 0xxx
Words 4-7: zeros
Word 8: 99 0000 0000

where pp = number of variables, kk = number of factors, xxx = identification.

e. Console set-up:

(1) Storage Entry Switches to 70 1951 1000
(2) Standard Settings otherwise

f. Input:
(1) Program Deck (6.6.021.5)
(2) Header Card
(3) F matrix cards (from 6.6.021.3 or from output of this program, discarding h^2 cards)
(4) Complete R matrix cards from output 6.6.020.1
(5) Trailer card (select appropriate one).

After trailer card has been read, (end of file key depressed) punch-out will occur (Trailer Cards 1 and 3) or not (Trailer Card 2). After the end of punch-out, machine will stop on Input-Output. Start with step 4 above to obtain next iteration, i.e., load R matrix and trailer card, depress Read Start key, depress Program Start key. If any error occurs, read Trailer Card 1 and obtain punch-out of F at this stage.

g. Stops: 01 0001 8000 Data card error, check columns 71-80. The F matrix should contain zeros, the header card should contain 9900000000.

g. Stops:

01 002 1500	End of F matrix punch-out. To continue program, reload the R matrix, depress Program Start and Start Read keys
01 0011 8000	Overflow in calculation of h^2, see General Flow
01 0012 8000	Overflow in $F_h{}'F$ matrix, see General Flow
01 0013 8000	Overflow in calculation of R_iF_j, see General Flow
01 0014 8000	Overflow in calculation of h^2 for new F_i from Doolittle solution, see General Flow
70 1951 1548	End of R deck. To continue the program, reload R and depress the Start Read key

Divide overflow indicates an h^2 greater than 0.999900, see General Flow

h. Output: (1) F_I matrix for use in 6.6.021.4 and for iterating 6.6.021.5, having the same format as the output of 6.6.021.3

(2) h^2 (communality) cards, punched 7 words per card, in order of variables, with card number in col. 76

Approx. time required: about 1½ hours for maximum sized and 20 minutes for medium sized problems, per iteration.

Title: SIMPLE STRUCTURE ROTATION

(Factor Analysis)

Description: The program receives an arbitrary matrix F(factor loadings), for example the output of the Centroid programs of Texas A & M's AM 557 or AM 561, or the output of the maximum-likelihood program (6.6.021.1, 6.6.021.3, 6.6.021.5 in this report) and identifies potential planes for a Simple Structure. For this purpose it starts with each variable or test as a trial and attempts to find a transformation vector which produces a plane with as many as possible "zero loadings" (values between ± .099) while allowing the trial vector to retain a high loading, if such a plane exists.

Method: The method used produces an oblique structure, i.e., the planes obtained are in no way forced to be orthogonal to each other, so that the artificial assumption of uncorrelated factors need not be made. It is based upon a combination of a weighted (but near singular) least-squares solution known as Analytical Method of Rotation,* and other rotational methods into a method known as Square-Root Weight Method[**].

---

* Thurstone, L.L., Analytical Method for Simple Structure, The Psychometric Laboratory, University of North Carolina, No. 6, July, 1953.

The termini of the test vectors in the common factor space are projected into k different planes (k = number of factors) and weights are determined on the basis of the distance of these projections from the trial vector. Then a new vector is found by weighted least-squares. Points at a certain distance (greater than ± .60 in the first iteration) are assigned weight zero. In subsequent iterations (four in all) these weights consider only points which lie much closer to the plane (usually ± .20 in second, ± .15 in third, and ± .10 in final, with built-in modifications, however). The program will usually, but not always, proceed to the final iteration unless, by such "scanning", no plane can be found starting with a particular trial.

Especially for very small studies (fewer than 10 variables) and large communalities it happens occasionally that fewer than k (number of factors) projections on the trial vector are below .60; if this occurs, the program will change the cut-off constant to .70, finally to .80; it will not proceed to the iterative procedure unless at least k projections are within the range.

[**] Bargmann, R.E., "A Comparison of New Analytical Methods for the Determination of Simple Structure." Hochschule fuer Internationale Paedagogische Forschung, Frankfurt/Main, Germany, September 1953.

Occasionally, therefore, the output vector for a particular trial will not be a simple structure plane, thus indicating that no such plane can be found, starting with that particular trial.

Limitations: The program is limited to a maximum of 35 variables (tests) and 10 factors.

Input:

1. Program Deck
2. Transfer card to 0000
3. Data cards
4. Trailer Card

Transfer card: Word 1: $\overset{12}{0}00000000\overset{12}{0}$

Words 2-8: arbitrary or blank.

Data cards: Word 1: Value of the factor loading in (4;6) form,i.e., with six decimals. The first four digits must,of course, be zero, since the factor loadings are correlations.

Word 2: 00 00aa 0000

where aa is the number of the variable (01, 02, 03 ... P); (row number of the element in word 1 in the F matrix).

Word 3: 00 00bb 0000

where bb is the number of the factor (01, 02 ... k); (column number of the element in word 1 in the F matrix).

Words 4 to 8: Zeros

These data cards have the output format of AM 557 or AM 561 (centroid program) or of the maximum-likelihood program 6.6.021.5 Note, however, that prior to the output for each factor in AM 557 or AM 561 there is a card having zero in columns 15-16; these indicate the per cent of factoring and must be eliminated before the data deck is used for the present program.

Trailer card (load card):

Word 1: 00 0000 1650

Word 2: 00 00pp 0000

Word 3: 00 00kk 0000

Word 4: xx xxx0 0000

Words 5 to 8: blank.

High (12) punches in 1, 10, 20, 30, 40.

Where pp = number of variables (tests); (maximum 35);

kk = number of factors (maximum 10);

xxxxx = arbitrary numerical identification.

Console Setting: 70 1951 xxxx

where xxxx is any invalid address.

Switches: Standard, except that OVERFLOW is set to SENSE.

Programmed Stops: 0000 indicates that study is completed. A new study (new set of data cards) may be read in, consisting only of steps 3 and 4 of Input above. Depress PROGR START on console and READ START on the 533. Do not depress COMP RESET.

Output:

Word 1: xx xxxy aacc

where xxxxx = Identification from Input Trailer Card

y = 8, if words 2 to 8 contain the elements of the transformation vector (vector by which F has been multiplied to obtain the plane)

= 9, if words 2 to 8 contain the loadings
on the rotated plane.

aa = Number of the variable which was used
as the trial

cc = Card number (1 or 2 for y = 8-cards,
from 1 to 5 for y = 9-cards.)

Words 2-8: loadings in (4;6) form in order.

Program Description

After the data cards have been read in, the program will start in location 1650. It will take variable (test) no. 1 as trial vector and attempt to find a simple structure plane with this one as a trial. After completion it will punch out the result and proceed taking variable (test) no. 2, etc., until each variable (test) has been used as a trial.

Unprogrammed Stops and Interruptions

In case of any unprogrammed stop, especially if overflow occurs with 1580010277 in PROGRAM REGISTER, Transfer, manually to 0117, Depress COMP RESET, Depress PROGR START.

Comment: One such condition arose in one of the many studies which were tried out; the sum-of-squares of corrected values for the transformation vector exceeded 10. The correction, in this case, would be so excessively large as to be of little value. The above procedure bypasses this particular trial vector.

If not every variable is desired as a trial vector, particularly if a very long run has been interrupted, proceed as follows:

1. Address-Stop at 0029
2. Read into storage, from console switches, in data address, into location 0028, the number of the last completed trial or the next desired trial minus one. For example, if run was interrupted after trial vector 12 was used, start program as if you were to start anew (Transfer card to zero and data cards for F if program is still loaded) address-stop at 0029, and read

   00 0012 0000 + via console switches into location 0028.

   Or, if during the running you want to skip to variable 12, say, address-stop at 0029, read

   00 0011 0000 via console switches into location 0028.
3. Then transfer manually back to 0029, i.e.,

Set CONTROL SWITCH to MANUAL

Set DISPLAY SWITCH to ANYTHING EXCEPT READ-IN STOR.

Set ADDRESS SELECTION SWITCHES to 0029

Depress COMP RESET; Depress TRANSFER

Set CONTROL SWITCH TO RUN

Depress PROGRAM START

Timing:

For one plane after a complete set of iterations (4)

$25 + 1.1 \times p \times k$ sec.

In normal operation (without skipping trials) multiply this by p in order to obtain the maximum time required for the entire program. For example, with 25 variables ($p = 25$) 10 factors ($k = 10$)

$t = 25 + 1.1 \times 25 \times 10 = 300$ sec = 5 min,

(observed average was 29¼ sec per plane)

Hence total run takes no longer than $5 \times 25 = 125$ min.

LAWLEY'S SOLUTION OR PRINCIPAL AXES

6.6.021.7 Obtain $F_h'F$ matrix or $F'F$ matrix, where F_h' stands for $F'D_{1/1-h^2}$.

a. Purpose: using the final output (last iteration) from 6.6.021.5, obtain the matrix $F_h'F$ or $F'F$, ready for input of eigenvector routine 5.2.005.2.

b. Program: 6.6.021.7

c. 533 Control Panel: 80 - 80 board with these additions:

(1) Jackplug P +
(2) Plug DI to C on Punch Digit Emitter
(3) Wire 12 from Punch Digit Emitter to 12-X of Punch Column Split 1
(4) Wire highest order digit of word 1 of Storage Exit C (the one that is connected with column 1 of Punch Card C on the standard board) to "Normal" of Punch Column Split 1
(5) Wire "Common" of Punch Column Split 1 to col. 1 of Punch Card C.

d. Header Card: Same as the one used in 6.6.021.5

e. Console Set-up:

(1) Storage Entry to 70 1951 1000

(2) Standard Settings otherwise

f. Input:

(1) Program deck 6.6.021.5

(2) Optional card for obtaining F'F if Principal Axes are desired; omit this step if Lawley's Solution is desired.

(3) Header Card

(4) F matrix, final output of 6.6.021.5; discard communality cards (last 5 cards in output).

(5) Program deck 6.6.021.7

(6) Trailer Card

g. Stops:

01 0C09 8000 End

h. Output: F_h'F or F'F in load card form, for input to 5.2.005.2

Approx. time required: 10 mins maximum.

5.2.005.2 and 6.6.021.8 Obtain E matrix

a. Purpose: to obtain the characteristic roots and eigenvectors of these matrices

b. Programs: Library Routine 5.2.005.2 and 6.6.021.8

c. 533 Control Panel: 80 - 80 Board

d. Header card: load card with "12" punches in card col's 1, 10, 20, 30

Col's	Field
1-10	69 1952 1953
11-20	00 00(n)0000 where n is the order of F_h'F, i.e. n = number of factors
21-30	24 0011 8000

e. Console set-up:
(1) Storage Entry to 70 1951 1951
(2) Standard settings otherwise

f. Input:
(1) Lib. R.5.2.005.2 Program cards
(2) Transfer card (data card with 00 0000 8000)
(3) Header card

6.6.021.8 (cont.)

f. Input:

(4) $F_h'F$ or $F'F$ cards from 6.6.021.7

(5) Transfer card (data card containing 00 0000 0350 in card col's 1-10)

(6) Transfer card (data card containing 00 0000 8000 in card col's 1-10)

(7) 6.6.021.8 Program cards

g. Stops: 01 0100 0000 End

h. Output:

(1) Characteristic root matrix, identified by zero in col. 80, punched row-wise in 7 word per card form, for informational purposes only.

(2) Eigenvector (E) matrix for input to 6.6.021.9 punched row-wise.

Approx. time required: 15 mins maximum.

6.6.021.9 Obtain Lawley's or Principal Axes Solution.

a. Purpose: to obtain Lawley's form or the Principal Axes form of the maximum-likelihood solution.

b. Program: 6.6.021.9

c. Control Panel: 80 - 80 Board

d. Header Card: same one used in 6.6.021.5 and 7.

e. Console set up:
  (1) Storage Entry to 70 1951 1900
  (2) Standard Settings otherwise

f. Input
  (1) Program deck 6.6.021.9
  (2) Header Card
  (3) F matrix, output of 6.6.021.5, as used in 6.6.021.7
  (4) E matrix, output of 6.6.021.8
  (5) Trailer card (transfer to 1700)

g. Stops: 01 0001 8000 End

h. Output: Matrix FE = Lawley's Solution or Principal Axes Solution, depending on input step (2) of 6.6.021.7. Format is the same as that in "Complete R" in 6.6.020.1, i.e., the matrix is punched row-wise, 1 card per row if $k \leq 7$ or 2 cards per row if $k > 7$.

Approx. time required: 8 min maximum.